(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号
特開2003－44080
（P2003－44080A）
(43)公開日　平成15年２月14日(2003.2.14)

(51)Int.Cl.'	識別記号	ＦＩ		テーマコード'（参考）
Ｇ１０Ｌ　15/06		Ｇ０６Ｋ　9/00	Ｓ	５Ｂ０６４
Ｇ０６Ｋ　9/00		Ｇ１０Ｌ　3/00	５２１Ｃ	５Ｄ０１５
Ｇ１０Ｌ　15/00			５２１Ｊ	
15/14			５３５Ｚ	
15/20			５５１Ｈ	

審査請求　未請求　請求項の数22　ＯＬ　（全 25 頁）　最終頁に続く

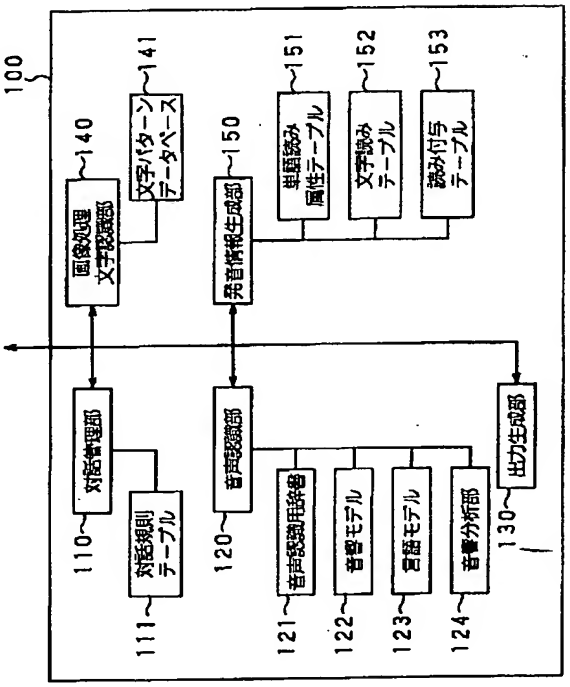
(21)出願番号　特願2002－130905(P2002－130905)
(22)出願日　平成14年５月２日(2002.5.2)
(31)優先権主張番号　特願2001－135423(P2001－135423)
(32)優先日　平成13年５月２日(2001.5.2)
(33)優先権主張国　日本（ＪＰ）

(71)出願人　000002185
ソニー株式会社
東京都品川区北品川６丁目７番35号
(72)発明者　廣江　厚夫
東京都品川区北品川６丁目７番35号　ソニー株式会社内
(72)発明者　南野　活樹
東京都品川区北品川６丁目７番35号　ソニー株式会社内
(74)代理人　100067736
弁理士　小池　晃　（外２名）

最終頁に続く

(54)【発明の名称】　ロボット装置、文字認識装置及び文字認識方法、並びに、制御プログラム及び記録媒体

(57)【要約】
【課題】　未登録の単語を新規単語として認識用辞書に登録する。
【解決手段】　ＣＣＤカメラ２０において撮像された画像の文字認識の結果から推定される複数の文字と、これら各文字から推定される複数の読み仮名と、各読み仮名に対応する読み方とを発音情報生成部１５０において生成し、ここで得られた複数の読み方とマイク２３において取得したユーザからの発声とをマッチングすることによって、生成された複数候補の中から１つの読み仮名及び発音のしかた（読み方）を特定する。



1

【特許請求の範囲】

【請求項１】　内部状態に応じて自律的に動作するロボット装置において、

単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、

単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、

被写体を撮像する撮像手段と、

上記撮像手段において撮像された画像から所定パターンの画像を抽出する画像認識手段と、

周囲の音を取得する集音手段と、

上記集音手段において取得された音から音声を認識する音声認識手段と、

上記画像認識手段において抽出された上記所定パターンから推定される複数通りの表音文字を上記単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、

上記発音情報生成手段において生成された各音声波形と上記音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を上記画像認識手段において抽出されたパターン認識結果に対応する発音のしかたであるとして上記音声認識用辞書に新規に記憶する記憶制御手段とを備えることを特徴とするロボット装置。

【請求項２】　上記所定パターンの画像は、文字及び／又は複数個の文字からなる文字列であることを特徴とする請求項１記載のロボット装置。

【請求項３】　上記画像から抽出される複数個の文字と該文字に対して付与される複数通りの発音のしかたとの対応を一時辞書として一時的に記憶する一時記憶手段を備えることを特徴とする請求項２記載のロボット装置。

【請求項４】　単語と該単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備え、上記記憶制御手段は、新規に記憶する文字と該文字の発音のしかたとともに上記単語属性を対応させて上記音声認識用辞書に記憶することを特徴とする請求項２記載のロボット装置。

【請求項５】　上記音声認識手段において認識された音声に対する応答を生成する対話管理手段を備え、上記対話管理手段は、上記単語属性を音声に対する応答規則で使用することを特徴とする請求項４記載のロボット装置。

【請求項６】　上記音声認識手段は、隠れマルコフモデル法に基づいて音声を認識することを特徴とする請求項２記載のロボット装置。

【請求項７】　単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、

単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、

2

被写体を撮像する撮像手段と、

上記撮像手段において撮像された画像から所定パターンの画像を抽出する画像認識手段と、

周囲の音を取得する集音手段と、

上記集音手段において取得された音から音声を認識する音声認識手段と、

上記画像認識手段において抽出された上記所定パターンの画像から推定される複数通りの表音文字を上記単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、

上記発音情報生成手段において生成された各音声波形と上記音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を上記抽出した文字の発音のしかたであるとして上記音声認識用辞書に新規に記憶する記憶制御手段とを備えることを特徴とする文字認識装置。

【請求項８】　上記所定パターンの画像は、文字及び／又は複数個の文字からなる文字列であることを特徴とする請求項７記載の文字認識装置。

【請求項９】　上記画像から抽出される複数個の文字と該文字に対して付与される複数通りの発音のしかたとの対応を一時辞書として一時的に記憶する一時記憶手段を備えることを特徴とする請求項７記載の文字認識装置。

【請求項１０】　単語と該単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備え、上記記憶制御手段は、新規に記憶する文字と該文字の発音のしかたとともに上記単語属性を対応させて上記音声認識用辞書に記憶することを特徴とする請求項７記載の文字認識装置。

【請求項１１】　上記音声認識手段において認識された音声に対する応答を生成する対話管理手段を備え、上記対話管理手段は、上記単語属性を音声に対する応答規則で使用することを特徴とする請求項１０記載の文字認識装置。

【請求項１２】　上記音声認識手段は、隠れマルコフモデル法に基づいて音声を認識することを特徴とする請求項７記載の文字認識装置。

【請求項１３】　被写体を撮像する撮像工程と、

上記撮像工程において撮像された画像から所定パターンの画像を抽出する画像認識工程と、

周囲の音を取得する集音工程と、

上記集音工程において取得された音から音声を認識する音声認識工程と、

上記画像認識工程において抽出された所定パターンの画像から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成工程と、

上記発音情報生成工程において生成された各音声波形と上記音声認識工程において認識された音声の音声波形とを比較し、最も近い音声波形を上記抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶制御工程とを備えることを特徴とする文字認識方法。

【請求項１４】　上記所定パターンの画像は、文字及び／又は複数個の文字からなる文字列であることを特徴とする請求項１３記載の文字認識方法。

【請求項１５】　上記画像から抽出される複数個の文字と該文字に対して付与される複数通りの発音のしかたとの対応を一時辞書として一時記憶手段に記憶する工程を備えることを特徴とする請求項１４記載の文字認識方法。

【請求項１６】　上記記憶制御工程では、新規に記憶する文字と該文字の発音のしかたとともに単語属性を対応させて上記音声認識用辞書に記憶することを特徴とする請求項１４記載の文字認識方法。

【請求項１７】　上記音声認識工程において認識された音声に対する応答を生成する対話管理工程を備え、上記対話管理工程では、上記単語属性が音声に対する応答規則で使用されることを特徴とする請求項１６記載の文字認識方法。

【請求項１８】　上記音声認識工程では、隠れマルコフモデル法に基づいて音声認識されることを特徴とする請求項１４記載の文字認識方法。

【請求項１９】　内部状態に応じて自律的に動作するロボット装置の制御プログラムにおいて、
被写体を撮像する撮像処理と、
上記撮像処理によって撮像された画像から所定パターンの画像を抽出する画像認識処理と、
周囲の音を取得する集音処理と、
上記集音処理によって取得された音から音声を認識する音声認識処理と、
上記画像認識処理によって抽出された所定パターンの画像から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成処理と、
上記発音情報生成処理によって生成された各音声波形と上記音声認識処理において認識された音声の音声波形とを比較し、最も近い音声波形を上記抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶処理とをロボット装置に実行させることを特徴とする制御プログラム。

【請求項２０】　上記所定パターンの画像は、文字及び／又は複数個の文字からなる文字列であることを特徴とする請求項１９記載の制御プログラム。

【請求項２１】　被写体を撮像する撮像処理と、
上記撮像処理によって撮像された画像から所定パターンの画像を抽出する画像認識処理と、
周囲の音を取得する集音処理と、
上記集音処理によって取得された音から音声を認識する音声認識処理と、
上記画像認識処理によって抽出された所定パターンの画像から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、上記付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成処理と、
上記発音情報生成処理によって生成された各音声波形と上記音声認識処理において認識された音声の音声波形とを比較し、最も近い音声波形を上記抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶処理とをロボット装置に実行させるための制御プログラムが記録された記録媒体。

【請求項２２】　上記所定パターンの画像は、文字及び／又は複数個の文字からなる文字列であることを特徴とする請求項２１記載の記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は、内部状態に応じて自律的に動作するロボット装置、文字認識装置及び文字認識方法、並びに、制御プログラム及び記録媒体に関し、特に、撮像した画像から所定パターンの画像を認識し、この画像とともに取得した音声をこの認識画像と対応付けて新規に登録するロボット装置、並びに、撮像された所定パターンの画像とともに取得した音声をこの認識画像と対応付けて新規に登録する文字認識装置及び文字認識方法、並びに、取得した画像から所定パターンの画像を認識し、この画像とともに取得した音声をこの認識画像と対応付けて新規に登録する処理を実行させる制御プログラム及びこの制御プログラムが記録された記録媒体に関する。

【０００２】

【従来の技術】電気的又は磁気的な作用を用いて人間（生物）の動作に似た運動を行う機械装置を「ロボット」という。我が国においてロボットが普及し始めたのは、１９６０年代末からであるが、その多くは、工場における生産作業の自動化・無人化等を目的としたマニピュレータや搬送ロボット等の産業用ロボット（Industrial Robot）であった。

【０００３】最近では、人間のパートナーとして生活を支援する、すなわち住環境その他の日常生活上の様々な場面における人的活動を支援する実用ロボットの開発が進められている。このような実用ロボットは、産業用ロボットとは異なり、人間の生活環境の様々な局面におい

て、個々に個性の相違した人間、又は様々な環境への適応方法を自ら学習する能力を備えている。例えば、犬、猫のように４足歩行の動物の身体メカニズムやその動作を模した「ペット型」ロボット、或いは、２足直立歩行を行う動物の身体メカニズムや動作をモデルにしてデザインされた「人間型」又は「人間形」ロボット（Humanoid Robot）等の脚式移動ロボットは、既に実用化されつつある。これらの脚式移動ロボットは、動物や人間の容姿にできるだけ近い外観形状を有し、産業用ロボットと比較して動物や人間の動作に近い動作を行うことができ、更にエンターテインメント性を重視した様々な動作を行うことができるため、エンターテインメントロボットと呼称される場合もある。

【０００４】脚式移動ロボットの中には、「目」に相当する小型カメラや、「耳」に相当する集音マイク等を備えているものもある。この場合、脚式移動ロボットは、取得した画像に対して画像処理を施すことによって、画像情報として入力した周囲の環境を認識したり、入力した周囲の音から「言語」を認識したりできる。

【０００５】特に、外部から取得した音声を認識して文字に変換したり、音声を認識して応答したりする手法は、脚式移動ロボット以外にもパーソナルコンピュータや、その他の電子機器に音声認識装置として適用されている。

【０００６】従来の音声認識の手法では、単語の発音と表記とが対応付けされて記憶された音声認識用の辞書（以下、認識用辞書と記す。）を用いて音声認識している。そのため、認識用辞書に登録されていない単語に関しては認識できないという欠点があった。更に、「文」のような連続した単語の発音を認識する場合には、認識用辞書に登録されている単語の組み合わせでなくてはならない。つまり、認証用辞書に登録されていない単語が含まれる場合、誤認識されるか、認識できない。

【０００７】「北品川」という単語を例にとると、「北品川」が認証用辞書に登録されていなければ、「北品川」及び「北品川」を含む発音、例えば、「北品川は、どこですか。」という単語の連続からなる音声は、認識できないか、「北品川」の部分が誤認識される。そこで、認識用辞書に登録されていない単語を認識できるようにするためには、未登録の単語を新たに追加登録することが必要になる。

【０００８】音声認識装置が音声認識を可能とするために備える認識用辞書とは、他の単語と区別するための識別子としての「単語シンボル」と、その単語の発音情報を表す「ＰＬＵ列」とが対応付けられたものである。ＰＬＵ（Phonone-like unit）とは、音響的及び音韻的単位となるものである。発音された音声は、ＰＬＵの組み合わせ（ＰＬＵ列）として必ず表現することができる。

【０００９】したがって、認識用辞書に単語を登録する場合は、単語シンボルとこれに対応するＰＬＵ列とを追加すればよい。ただし、単語シンボルとＰＬＵ列とを追加できる場合というのは、「北品川」や「ｋｉｔａｓｈｉｎａｇａｗａ」という表記を、例えば、キーボード等のような入力手段を用いて直接入力できる場合に限られる。

【００１０】そのため、ロボット装置のようにキーボードのような入力手段を備えていない場合には、音声として取得した単語の発音を音声認識して未知単語のＰＬＵ列を得る方法もある。この場合、ガーベージモデル（garbage model）を適用して認識している。ガーベージモデルとは、図２０（ａ）及び図２０（ｂ）に示すように、音声を発音の基本的な単位となる「音素」の組み合わせとして表した、また、単語の読み方の基本的な単位となる「かな」の組み合わせとして表した（ただし、日本語の場合。）モデルである。

【００１１】従来の音声認識装置では、ガーベージモデルを適用することによって、音声による認識結果が得て、この認識結果に単語シンボルを当てはめて、これらを対応させて新規単語として認識用辞書に登録している。

【００１２】ただし、ここで「音素」と「ＰＬＵ」とは、ほぼ同義の単語として使用しており、「ＰＬＵ列」は、複数の「ＰＬＵ」が接続されることで構成された単語の発音を表記したものである。

【００１３】

【発明が解決しようとする課題】ところが、ガーベージモデルを適用した従来の音声認識の手法では、同じ単語であってもユーザ毎に発声のしかたに微妙な違いがあることや、弱い音素（例えば、語頭の／ｓ／等）は、必然的に認識されにくくなることや、周囲の雑音の影響による音素の変化や、音声区間検出の失敗等が原因となって、認識精度が悪くなるという欠点があった。

【００１４】特に、ロボット装置に音声認識装置を適用した場合、音声認識装置側の音声取得用のマイクとユーザ（音声源）との距離が離れている状況下で使用されることが多いため、誤認識の頻度が高くなる。

【００１５】具体的に、例えば、「きたしながわ」を認識させる場合について示すと、認識結果は、「ｈｉ　ｔｏ　ｔｓｕ　ｎａ　ｎｏ　ｇａ」や「ｉ　ｔａｓ＿ｎａ　ｇａ：」のように「きたしながわ」と類似しているが、同一ではないＰＬＵ列として認識されることがある。このような方法で単語登録された辞書を用いて音声認識を行うと、認識精度の低下、また誤認識による表示誤り等の問題が発生する。つまり、新規登録語には、不正確なＰＬＵ列が付与されていることになるため、この単語を認識する際の精度が低下するという問題点があった。

【００１６】登録した人とは別の人が同じ単語を発音した場合、仮に「きたしながわ」が認識用辞書に登録されていたとしても、ユーザ毎の発音の癖から「きたしなが

わ」という単語を含む発音が認識されないこともあった。

【００１７】また、音声認識の結果を文字に変換して表示する場合、新規登録語には、表示に関する情報が与えられていないため、誤った文字が表示されることがある。ユーザが「きたしながわ」を音声で登録した後、音声認識装置に対して「北品川に行きたい。」と発声した場合、音声認識装置には「きたしながわ」が正しく認識されたとしても、表示は「ｈｉｔｏｔｓｕｎａｎｏｇａに行きたい」や「『ひとつなのが』に行きたい」になることがある。また、音声認識装置が認識結果のＰＬＵ列を音声合成で反復する場合も、合成された新規登録語のＰＬＵ列の部分だけが不自然な繋がりとして発声されるという不都合も生じる。

【００１８】更に、このようにガーベージモデルによって登録された新規登録語は、品詞や意味等の単語の属性に関する情報を登録することができない。例えば、「北品川」を登録したとしても、この単語が名詞であるか地名であるかを表す情報を登録することができない。そのため、仮に、例えば、対話用の文法や認識用の言語モデル等に「＜地名を表す語＞＋は＋どこ＋です＋か」のような特定表現のための文法規則が予め記録されていたとしても、新規登録語には適用できないという問題点があった。登録時に単語の属性についても音声で入力することができるが、ユーザが単語の属性を知っている必要があった。また、単語の登録操作に加えて属性を入力することはユーザにとって煩わしい。

【００１９】そこで本発明は、このような従来の実情に鑑みて提案されたものであり、提示された文字とともに発音される音声に対して、撮像した画像から文字を認識し取得した音声をこの文字の発音として認識することによって、未登録の単語を新規単語として認識用辞書に登録でき、更に登録された新規単語を精度よく認識できるロボット装置、並びに、提示された文字とともに発音される音声に対して、撮像した画像から文字を認識し取得した音声をこの文字の発音として認識することによって、未登録の単語を新規単語として認識用辞書に登録でき、登録された新規単語を精度よく認識できる文字認識装置、及び、提示された文字を撮像し、撮像された画像から文字を認識し、提示とともに発音された音声を取得して認識された文字の発音として認識することによって、認識用辞書に新規単語として登録する文字認識方法、並びに、撮像した画像から文字を認識し取得した音声をこの文字の発音として新規に登録する処理を実行させる制御プログラム及びこの制御プログラムが記録された記録媒体を提供することを目的とする。

【００２０】

【課題を解決するための手段】上述した目的を達成するために、本発明に係るロボット装置は、単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、被写体を撮像する撮像手段と、撮像手段において撮像された画像から所定パターンの画像を抽出する画像認識手段と、周囲の音を取得する集音手段と、集音手段において取得された音から音声を認識する音声認識手段と、画像認識手段において抽出された所定パターンの画像から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する記憶制御手段とを備える。

【００２１】このようなロボット装置は、画像認識手段において抽出された所定パターンの画像から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成し、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した所定パターンの画像に対応する発音のしかたであるとして音声認識用辞書に新規に記憶する。

【００２２】ここで特に、所定パターンの画像は、文字及び／又は複数個の文字からなる文字列である。

【００２３】また、本発明に係る文字認識装置は、単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、被写体を撮像する撮像手段と、撮像手段において撮像された画像から文所定パターンの画像を抽出する画像認識手段と、周囲の音を取得する集音手段と、集音手段において取得された音から音声を認識する音声認識手段と、画像認識手段において抽出された所定パターンの画像から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する記憶制御手段とを備える。

【００２４】このような文字認識装置は、画像認識手段において抽出された所定パターンの画像から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音

のしかたと発音に相当する音声波形とを生成し、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する。

【００２５】ここで特に、所定パターンの画像は、文字及び／又は複数個の文字からなる文字列である。

【００２６】また、本発明に係る文字認識方法は、被写体を撮像する撮像工程と、撮像工程において撮像された画像から所定パターンの画像を抽出する画像認識工程と、周囲の音を取得する集音工程と、集音工程において取得された音から音声を認識する音声認識工程と、画像認識工程において抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成工程と、発音情報生成工程において生成された各音声波形と音声認識工程において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶制御工程とを備える。

【００２７】このような文字認識方法によれば、画像認識工程において抽出された所定パターンの画像から推定される複数通りの表音文字が単語表音テーブルに基づいて付与され、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形が生成され、発音情報生成工程において生成された各音声波形と音声認識工程において認識された音声の音声波形とが比較され、最も近い音声波形が抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶される。

【００２８】ここで特に、所定パターンの画像は、文字及び／又は複数個の文字からなる文字列である。

【００２９】更に、本発明に係る制御プログラムは、被写体を撮像する撮像処理と、撮像処理によって撮像された画像から所定パターンの画像を抽出する画像認識処理と、周囲の音を取得する集音処理と、集音処理によって取得された音から音声を認識する音声認識処理と、画像認識処理によって抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成処理と、発音情報生成処理によって生成された各音声波形と音声認識処理において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶処理とをロボット装置に実行させる。

【００３０】ここで特に、所定パターンの画像は、文字及び／又は複数個の文字からなる文字列である。また、上述の制御プログラムを記録媒体に記録して提供する。

【００３１】

【発明の実施の形態】本発明の一構成例として示すロボット装置は、内部状態に応じて自律動作するロボット装置である。このロボット装置は、少なくとも上肢と体幹部と下肢とを備え、上肢及び下肢、又は下肢のみを移動手段とする脚式移動ロボットである。脚式移動ロボットには、４足歩行の動物の身体メカニズムやその動きを模倣したペット型ロボットや、下肢のみを移動手段として使用する２足歩行の動物の身体メカニズムやその動きを模倣したロボット装置があるが、本実施の形態として示すロボット装置は、４足歩行タイプの脚式移動ロボットである。

【００３２】このロボット装置は、住環境その他の日常生活上の様々な場面における人的活動を支援する実用ロボットであり、内部状態（怒り、悲しみ、喜び、楽しみ等）に応じて行動できるほか、４足歩行の動物が行う基本的な動作を表出できるエンターテインメントロボットである。

【００３３】このロボット装置は、特に「犬」を模した形体であり、頭部、胴体部、上肢部、下肢部、尻尾部等を有している。各部の連結部分及び関節に相当する部位には、運動の自由度に応じた数のアクチュエータ及びポテンショメータが備えられており、制御部の制御によって目標とする動作を表出できる。

【００３４】このロボット装置は、周囲の状況を画像データとして取得するための撮像部、周囲の音声を取得するマイク部、外部から受ける作用を検出するための各種センサ等を備えている。撮像部としては、小型のＣＣＤ（Charge-Coupled Device）カメラを使用する。

【００３５】本実施の形態として示すロボット装置は、画像認識装置及び音声認識装置を備えており、ＣＣＤカメラにおいて撮像された画像から所定パターンの画像を抽出し、抽出された所定パターンの画像から推定される複数通りの読み仮名を付与し、付与された複数通りの読み仮名のそれぞれに相当する音声波形を生成する。ここでの画像の所定パターンとしては、文字（文字列）、物体の形状、輪郭、柄、物体そのものの画像等があげられる。そして、この音声波形とマイク部において取得した音声の音声波形とを比較し、最も近い音声波形を抽出した所定パターンの画像に対応する発音のしかた（読み方）であるとして音声認識用辞書に新規に記憶することができるロボット装置である。

【００３６】以下、本発明の一構成例として示すロボット装置について、図面を参照して説明する。以下の説明では、取得した画像から認識される所定パターンが文字（文字列）である場合について詳細に説明する。

【００３７】本実施の形態では、ロボット装置１は、図

１に示すように、「犬」を模した形状のいわゆるペット型ロボットである。ロボット装置１は、胴体部ユニット２の前後左右に脚部ユニット３Ａ、３Ｂ、３Ｃ、３Ｄが連結され、胴体部ユニット２の前端部に頭部ユニット４が連結され、後端部に尻尾部ユニット５が連結されて構成されている。

【００３８】胴体部ユニット２には、図２に示すように、ＣＰＵ（Central Processing Unit）１０、ＤＲＡＭ（Dynamic Random Access Memory）１１、フラッシュＲＯＭ（Read Only Memory）１２、ＰＣ（Personal Computer）カードインターフェイス回路１３及び信号処理回路１４が内部バス１５を介して相互に接続されることにより形成されたコントロール部１６と、このロボット装置１の動力源としてのバッテリ１７とが収納されている。また、胴体部ユニット２には、ロボット装置１の向きや動きの加速度を検出するための角速度センサ１８及び加速度センサ１９が収納されている。

【００３９】頭部ユニット４には、外部の状況を撮像するためのＣＣＤ（Charge Coupled Device）カメラ２０と、使用者からの「撫でる」や「叩く」といった物理的な働きかけにより受けた圧力を検出するためのタッチセンサ２１と、前方に位置する物体までの距離を測定するための距離センサ２２と、外部音を集音するためのマイク２３と、鳴き声等の音声を出力するためのスピーカ２４と、ロボット装置１の「目」に相当するＬＥＤ（Light Emitting Diode）（図示せず）等が所定位置にそれぞれ配置されている。ＣＣＤカメラ２０は、頭部ユニット４の向く方向にある被写体を所定の画角で撮像することができる。

【００４０】各脚部ユニット３Ａ～３Ｄの関節部分、各脚部ユニット３Ａ～３Ｄと胴体部ユニット２との連結部分、頭部ユニット４と胴体部ユニット２との連結部分、尻尾部ユニット５と尻尾５Ａとの連結部分には、自由度数分のアクチュエータ２５１～２５ｎ及びポテンショメータ２６１～２６ｎがそれぞれ配設されている。アクチュエータ２５１～２５ｎは、例えば、サーボモータを構成として有している。サーボモータの駆動により、脚部ユニット３Ａ～３Ｄが制御されて目標の姿勢、或いは動作に遷移する。

【００４１】これら角速度センサ１８、加速度センサ１９、タッチセンサ２１、距離センサ２２、マイク２３、スピーカ２４及び各ポテンショメータ２６１～２６ｎ等の各種センサ並びにＬＥＤ及び各アクチュエータ２５１～２５ｎは、それぞれ対応するハブ２７１～２７ｎを介してコントロール部１６の信号処理回路１４と接続され、ＣＣＤカメラ２０及びバッテリ１７は、それぞれ信号処理回路１４と直接接続されている。

【００４２】信号処理回路１４は、上述の各センサから供給されるセンサデータや画像データ及び音声データを順次取り込み、これらをそれぞれ内部バス１５を介してＤＲＡＭ１１内の所定位置に順次格納する。また信号処理回路１４は、これとともにバッテリ１７から供給されるバッテリ残量を表すバッテリ残量データを順次取り込み、これをＤＲＡＭ１１内の所定位置に格納する。

【００４３】このようにしてＤＲＡＭ１１に格納された各センサデータ、画像データ、音声データ及びバッテリ残量データは、ＣＰＵ１０が当該ロボット装置１の動作制御を行う際に使用される。

【００４４】ＣＰＵ１０は、ロボット装置１の電源が投入された初期時において、フラッシュＲＯＭ１２に格納された制御プログラムを読み出して、ＤＲＡＭ１１に格納する。又は、ＣＰＵ１０は、図１に図示しない胴体部ユニット２のＰＣカードスロットに装着された半導体メモリ装置、例えば、いわゆるメモリカード２８に格納された制御プログラムをＰＣカードインターフェイス回路１３を介して読み出してＤＲＡＭ１１に格納する。

【００４５】ＣＰＵ１０は、上述のように信号処理回路１４よりＤＲＡＭ１１に順次格納される各センサデータ、画像データ、音声データ、及びバッテリ残量データに基づいて自己及び周囲の状況や、使用者からの指示及び働きかけの有無を判断している。

【００４６】更に、ＣＰＵ１０は、この判断結果とＤＲＡＭ１１に格納した制御プログラムとに基づく行動を決定する。ＣＰＵ１０は、当該決定結果に基づいてアクチュエータ２５１～２５ｎの中から必要とするアクチュエータを駆動することによって、例えば、頭部ユニット４を上下左右に動かしたり、尻尾部ユニット５の尻尾を動かしたり、各脚部ユニット３Ａ乃至３Ｄを駆動して歩行させたりする。また、ＣＰＵ１０は、必要に応じて音声データを生成し、信号処理回路１４を介してスピーカ２４に供給する。また、ＣＰＵ１０は、上述のＬＥＤの点灯・消灯を指示する信号を生成し、ＬＥＤを点灯したり消灯したりする。

【００４７】また、ＣＰＵ１０は、上述のようにロボットを自律的に制御するほかに、後述する対話管理部１１０等からの要求に応じてロボットを動作させる。

【００４８】これらの基本的な構成によって、ロボット装置１は、自己及び周囲の状況や、使用者からの指示及び働きかけに応じて自律的に行動する。

【００４９】更に、ロボット装置１は、認識した発音と認識した文字との対応を新規登録語として音声認識用辞書に登録するための構成として、胴体部ユニット２のコントロール部１６に、画像音声認識部１００を備えている。画像音声認識部１００は、図３に示すように、対話管理部１１０と、音声認識部１２０と、出力生成部１３０と、画像処理文字認識部１４０と、発音情報生成部１５０とを有している。音声認識用辞書とは、図４に示すように、他の単語と区別するための識別子としての「単語シンボル」と、この単語に対応する発音情報を表す「ＰＬＵ列」とを記録したテーブルである。この辞書を

参照することによって、単語の発音のしかた（読み方）、又は、発音に対応する単語の表記が抽出できる。

【００５０】具体的に、対話管理部１１０は、マイク２３から入力したユーザの発話、対話履歴等から入力した音声に対する応答を生成する。対話管理部１１０は、対話規則テーブル１１１に記憶された種々の対話規則に基づいて、入力した音声に対する応答パターンを生成する。

【００５１】音声認識部１２０は、ユーザの発話を対話管理部１１０で処理できる形式、例えば、テキスト形式、構文解析、対話用フレーム等に変換する。音声認識部１２０は、具体的には、音声認識用辞書１２１、音響モデル１２２、言語モデル１２３、音響分析部１２４等からなる。音響分析部１２４では、認識に必要な特徴量の抽出が微少時間間隔で行われる。例えば、得られた音声信号のエネルギ、零交差数、ピッチ、周波数特性、及びこれらの変化量等が抽出される。周波数分析には、線形予測分析（ＬＰＣ）、高速フーリエ変換（ＦＦＴ）、バンドパスフィルタ（ＢＰＦ）等が用いられる。

【００５２】音声認識部１２０は、音響モデル１２２と言語モデル１２３とを用いて、音響分析部１２４で生成された特徴量系列に対応する単語系列を決定する。認識手法としては、例えば、隠れマルコフモデル（Hidden Markov Model：以下、ＨＭＭと記す。）等が用いられる。

【００５３】ＨＭＭとは、状態遷移確率と確率密度関数とをもつ状態遷移モデルであり、状態を遷移しながら特徴量系列を出力する確率値を累積して尤度を決定する。その尤度の値を「スコア」として音声認識用辞書に記憶されている単語の発音のしかたと後述する画像処理文字認識部において認識された文字に対して付与される発音のしかたとのマッチングに使用する手法である。ＨＭＭの遷移確率及び確率密度関数等は、学習用データに基づく学習過程を通じて、予め学習して用意される値である。

【００５４】音響モデルは、音素（ＰＬＵ）、音節、単語、フレーズ、文等、それぞれの単位毎に用意することができる。例えば、日本語の仮名『あ』・『い』・『う』・『え』・『お』・『か』・『き』…『ん』を単位とする音響モデルを用いた場合、これらを組み合わせて接続することによって、『はい』、『いいえ』、『おはよう』、『いまなんじですか』等の言葉が構成できる。音素とは、単語の発音情報を表すものであり、音響的及び音韻的単位である。本明細書では、音素とＰＬＵ（Phonone-like unit）とを区別しないで使用している。発音された音声は、音素（ＰＬＵ）の組み合わせ（ＰＬＵ列）として必ず表現することができる。

【００５５】ＨＭＭによれば、このように構成された言葉とマイク２３において取得した音声の特徴量系列との類似度をスコアとして計算することができる。音響モデルから「言葉」を構成するための情報として、言語モデル１２３と音声認識用辞書１２１とが利用される。音声認識用辞書１２１とは、認識対象となる各単語を構成するための音響モデル（ここでは、仮名の一文字『あ』、『い』、・・・等を示す。）の接続のしかたを対応テーブルとして示した辞書であり、言語モデル１２３とは、単語と単語との接続のしかたの規則を示したものである。

【００５６】以下に示す例では、「単語」とは、認識処理の上で発音する際に、１つの纏まりとして扱う方が都合がよい単位のことを示しており、言語学的な単語とは必ずしも一致しない。例えば、以下の例では「北品川」を一単語として扱う場合があるが、これを「北」「品川」という２単語として扱っても構わない。更に、「北品川駅」や「北品川駅はどこですか」を発音する上での一単語として扱うこともできる。

【００５７】また、本明細書では、「読み仮名」とは、漢字、英単語の読み方を表記したひらがな又はカタカナの意として用い、「発音のしかた」とは、読み仮名の実際の発音をローマ字、又はローマ字と記号とを使用して表記したものであり、言語学的における「音素記号」に相当する。

【００５８】例えば、『～時から、～時まで』という文を扱う場合について考える。この場合、まず、『０（ぜろ）』『１（いち）』・・・『２４（にじゅうよん）』という単語と、『時（じ）』・『から』・『まで』という言葉のそれぞれに関して、音響モデル１２２を参照することによって、単語の接続のしかたが決定される。

【００５９】次に、『（数字を表す単語）』、『時』、『から』、『（数字を表す単語）』、『時』、『まで』という各単語を言語モデル１２３を参照することによって、文を構成するための各単語の接続のしかたが決定される。

【００６０】この音声認識用辞書１２１と言語モデル１２３とを用いてＨＭＭを適用することによって、『１時から２時まで』や『２時から５時まで』等の文と入力される特徴量系列との類似度がスコアとして計算できる。その中で最も高いスコアを有する単語系列からなる文を音声認識結果として出力する。

【００６１】音声認識処理におけるスコアの計算は、音響モデル１２２によって与えられる音響的なスコアと、言語モデル１２３によって与えられる言語的なスコアとを総合評価して行われる場合もある。

【００６２】言語的なスコアとは、例えば、連続するｎ個の単語間の遷移確率、又は連鎖確率に基づいて与えられるスコアである。遷移確率は、予め、大量のテキストから統計的に求められた値であり、ここでは、この遷移確率を「ｎグラム」と呼称する。

【００６３】なお、言語モデルは、文法やｎグラム中に単語を直接記述する以外にも、単語のクラス（単語をあ

る基準や属性にしたがって分類したもの）を記述する場合もある。

【００６４】例えば、地名を表す単語を集め、それに＜地名＞というクラス名称を与えた場合に「＜地名＞＋は＋どこ＋です＋か」という文法を記述したり、ｎグラム中に「＜地名＞＋は＋どこ」の遷移確率を用意しておくこともできる。この場合、ｎ＝３であり、正確には、遷移確率は、Ｐ（＜地名＞｜は、どこ｜）である。

【００６５】出力生成部１３０は、対話管理部１１０が生成した応答パターンを実際の動作に変換する。例えば、対話管理部１１０が「首を左右に振る＋『いいえ』と発声する」という応答パターンを生成した場合、出力生成部１３０は、これを受けて「首を左右に振る」に対応する動作パターンを生成しＣＰＵ１０に送るとともに、「いいえ」に対応する音声波形を生成しスピーカ２４から出力する。

【００６６】画像処理文字認識部１４０は、ＣＣＤカメラ２０で取り込んだ画像に含まれる文字列を文字パターンデータベース１４１に基づいて識別する。文字パターンデータベース１４１には、ひらがな、カタカナ、漢字、アルファベット、記号類、必要に応じて各国語の文字等の画像パターンが格納されている。画像処理文字識別部１４０は、ＣＣＤカメラ２０からの入力画像と文字パターンデータベース１４１に格納されている画像パターンとの間でマッチングを行い、入力画像に含まれている文字列を認識する。

【００６７】発音情報生成部１５０は、画像処理文字認識部１４０で認識された文字列に対応する発音情報、つまり文字列の読み仮名を生成し、更にその発音のしかた（読み方）を生成する。例えば、入力画像から「北品川」という文字列が認識された場合、「きたしながわ」という読み仮名を生成し、ＰＬＵ列で「ｋｉｔａｓｈｉｎａｇａｗａ」という発音のしかた（読み方）を生成する。

【００６８】単語読み属性テーブル１５１は、図４に示すように、単語（文字列）と読み仮名と属性の組を記述したテーブルである。属性とは、「地名」、「名前」、「動物」のように単語のもつ意味を示している。

【００６９】画像処理文字認識部１４０で認識された文字列がこのテーブルに含まれている場合は、このテーブルから読み仮名を抽出することで、読み仮名からその文字列の発音のしかた（読み方）を確定できる。単語読み属性テーブル１５１は、音声認識用辞書１２１とは独立に用意する。

【００７０】認識用辞書の語彙数には、認識速度や精度や処理上の都合で上限がある（例えば６万５５３６語）が、単語読み属性テーブル１５１にはそれらの制限とは関係なく単語を記述することができる。この単語読み属性テーブル１５１は、他の言語資源から流用することも可能である。例えば、仮名漢字変換プログラムや形態素解析プログラム等で使用されている辞書等を流用することもできる。

【００７１】文字読みテーブル１５２は、図６に示すように、文字と読み仮名との対応が記述されたテーブルである。記号やアルファベットや単漢字毎に読み仮名を記述しておく。使用可能な文字全てについて読み仮名を記述しておけば、任意の文字列に対して読み仮名から発音のしかた（読み方）を付与することができる。

【００７２】読み付与テーブル１５３は、２つのテーブルだけでは読み仮名が付与できない場合に読み仮名を付与するための規則や、読み仮名が特定できない場合に、これを特定するための規則が記述してある。例えば、音読み及び訓読みの統一、長音化に関する規則、連濁の規則、繰り返し記号に関する規則、英単語に読みを付与する規則がある。

【００７３】具体的には、長音化に関する規則とは、「・・・おう」「・・・えい」等を「・・・おー」「・・・えー」等に変換する規則である。この規則によって、例えば、「とうきょう」は、「とーきょー」に変換される。連濁の規則とは、例えば、「品川口」の読みを「しながわ（品川）」と「くち（口）」との結合から生成する場合に、「くち」を濁らせて「ぐち」にする規則である。また、繰り返し記号に関する規則とは、「々・ゝ・ゞ・ゝ・ゞ」等の繰り返し記号に対応して読み仮名を付ける規則である。更に、英単語に読み仮名を付与する規則とは、英単語の語末に“ｅ”がある場合は、“ｅ”自体は、発音しないかわりに前の母音を母音読みする等の規則である。例えば、“take”に「テーク」という読み仮名を付与する際に、“ａ”に対して「エー」という読み仮名を付与し、“ke”に対して、単に「ク」という読み仮名を付与する規則である。

【００７４】次に、認識用辞書に新規単語を登録する際の処理を、図７を用いて具体的に説明する。

【００７５】まず、ステップＳ１において、単語登録のための単語登録モードに移行する。単語登録モードへの移行は、例えば、ロボット装置１は、ユーザが発する「登録モード」や「言葉を覚えて」等の言葉をトリガとして単語登録モードに移行する。このほかに、操作ボタンを設け、この操作ボタンが押されたときに単語登録モードへ移行するようにしてもよい。

【００７６】ステップＳ２において、ロボット装置１は、ユーザに対して、登録したい単語の表記をロボット装置１のＣＣＤカメラ２０の前に提示する旨の指示及び／又は提示に加えてユーザが登録したい単語の読み方を発声する旨の指示を促す。ユーザに対する指示は、ロボット装置１が音声によって指示してもよいし、また、図示しない表示部に指示内容を表示する場合でもよい。ここでは、「北品川」という単語を例として説明する。ユーザによって提示される文字は、漢字でも仮名でもローマ字表記でもＰＬＵ列でも構わない。具体的には、ロボ

ット装置１は、「北品川」、「きたしながわ」、「キタシナガワ」、「ｋｉｔａｓｈｉｎａｇａｗａ」等の何れの表記も認識できる。

【００７７】ステップＳ３において、ロボット装置１は、文字提示のみであるか、文字提示とともに発話があったかを判断する。文字提示だけの場合は、ステップＳ４へ進み、文字提示とともに発話があった場合は、後述するステップＳ８へと進む。それ以外、すなわち、発声のみの場合は、従来と同様にガーベージモデルによる認識処理を行う。

【００７８】はじめに、文字提示のみの場合について説明する。文字提示のみの場合、ステップＳ４において、ロボット装置１における画像処理文字認識部１４０は、ＣＣＤカメラ２０において撮像された画像にどのような文字列が含まれているかを文字パターンデータベース１４１に基づいて、文字認識（ＯＣＲ：Optical Character Recognition）する。ここで、画像処理文字認識部１４０は、文字認識結果の候補が１つに絞り込めない場合、複数の候補を残す。例えば、「北品川」という文字に対して「比晶川」という認識結果が得られた場合は、「比晶川」も残す。

【００７９】続いて、ステップＳ５において、ロボット装置１における発音情報生成部１５０は、ステップＳ４での認識結果として得られた文字列に対して、文字列の発音のしかた（読み方）を生成する。発音を生成する際の詳細は、後述する。発音生成処理によって、文字列に対して発音のしかた（読み方）が付与される。認識された文字列が複数ある場合及び／又は１つの文字列に対して複数の発音のしかたが有り得る場合には、全ての発音パターンが適用される。

【００８０】ステップＳ６において、ロボット装置１は、上述のように生成された文字列に対する発音のしかた（読み方）が正しいか否か、又は、複数の読み方のうちどれを採用すべきかをユーザに確認する。発音のしかた（読み方）が一通りのみの場合は、「読み方は、○○で正しいですか。」のように質問する。ユーザが「正しい」や「はい」等の応答を返した場合は、ステップＳ７に進む。

【００８１】また、発音のしかた（読み方）が複数通りある場合は、それぞれについて「読み方は、○○ですか。」のように質問する。ユーザが「正しい」や「はい」等の応答を返した読み方を採用してステップＳ７に進む。

【００８２】ユーザから「いいえ」等の応答を受けた場合、すなわち、正しい読み方が存在しない場合、ステップＳ２若しくはステップＳ４の処理まで戻る。

【００８３】以上の処理によって、新規単語の読みを確定した後、ステップＳ７に進み、取得した文字列とこの文字列に対する発音のしかた（読み方）とを対応付けて新規単語として認識用辞書に登録する。新規単語を追加する際、図４に示す単語シンボル欄には、提示された文字の認識結果を使用する。この文字列に対応するＰＬＵ列欄には、ステップＳ６において確定した発音のしかた（読み方）が記述される。新規単語を登録した後、登録モードを終了する。その後、更新された認識用辞書を音声認識に反映させるための処理、例えば、音声認識プログラムの再起動等を行う。

【００８４】一方、ステップＳ３において、ユーザが文字を提示するとともに表記した文字を発声した場合について説明する。文字提示ともに発話があった場合は、両者から得られる情報を協調的に使用することによってＰＬＵ列等の発音情報を精度よく生成することができる。

【００８５】具体的には、文字認識の結果から推定される複数の文字と、これら各文字から推定される複数の読み仮名と、各読み仮名に対応する発音のしかた（読み方）とを生成する。このようにして得られた複数の発音のしかた（読み方）とマイク２３において取得したユーザからの発声とをマッチングすることによって、上述のように生成された複数候補の中から１つの読み仮名及び発音のしかた（読み方）を特定する。

【００８６】文字提示とともに発話があった場合、ステップＳ８において、ロボット装置１における画像処理文字認識部１４０は、ＣＣＤカメラ２０において撮像された画像から文字認識する。ここで、画像処理文字認識部１４０は、文字認識結果の候補が１つに絞り込めない場合、複数の候補を残す。

【００８７】続いて、ステップＳ９において、ロボット装置１における発音情報生成部１５０は、ステップＳ８での認識結果として得られた文字列に対して、文字列の読み仮名を生成する。発音生成処理によって、文字列に対して発音のしかた（読み方）が付与される。認識された文字列が複数ある場合及び／又は１つの文字列に対して複数の読み方が可能な場合には、全ての発音パターンが適用される。

【００８８】次に、ステップＳ１０において、文字列と発音のしかた（読み方）とから、一時的に仮の認識用辞書を生成する。この辞書を以下、新規単語用認識用辞書と記す。例えば、ＣＣＤカメラ２０によって撮像された「北品川」という文字が画像処理文字認識部１４０において、「北品川」と「比晶川」の２通りに認識されたとする。音声情報生成部１５０は、「北品川」と「比晶川」に読み仮名を付与する。「北品川」には「きたしながわ」が付与され、「比晶川」には「ひしょうがわ」と「くらあきらがわ」の２通りが付与され、更に両者の発音のしかた（読み方）、すなわち、ＰＬＵ列が生成される。この場合の新規単語用認識用辞書を図８に示す。

【００８９】ステップＳ１１において、新規単語用認識用辞書を用いて、ユーザからの発声に対して音声認識を行う。ここでの音声認識は、連続音声認識ではなく、単語音声認識である。新規単語用認識用辞書が生成される

よりも前にユーザが発話している場合は、その発話を録音しておき、その録音音声に対して音声認識を行う。ステップＳ１１における音声認識とは、新規単語用認識用辞書に登録されている単語の中からユーザの発話と音響的に最も近い単語を探し出すことである。ただし、ステップＳ１１の処理では、単語シンボルが同一であっても、ＰＬＵ列が異なる場合は別の単語とみなす。

【００９０】図８では、ここに登録されている３単語（２つの「比晶川」は別単語とみなす）の中から、ユーザの発話である「きたしながわ」に最も近い単語を探し出すことである。結果として、単語シンボルとＰＬＵ列との組を１つに特定することができる。

【００９１】新規単語用認識用辞書の中から単語シンボルとＰＬＵ列との組が特定されたら、ステップＳ７において、これを正規の音声認識用辞書１２１に登録する。新規単語を登録した後、登録モードを終了する。その後、更新された認識用辞書を音声認識に反映させるための処理、例えば、音声認識プログラムの再起動等を行う。

【００９２】以上示した処理によって、ロボット装置１は、音声認識用辞書１２１に記憶されていない単語を新規単語として登録できる。

【００９３】上述したステップＳ５とステップＳ９での文字列の発音のしかた（読み方）の生成に関して、図９を用いて詳細に説明する。

【００９４】まず、ステップＳ２１において、画像処理文字認識部１４０によって認識された文字列が仮名文字だけで構成されているか否かを調べる。ただし、ここでの仮名文字とは、ひらがな・カタカナのほかに長音記号「ー」や繰り返し記号「々…」等も含む。文字列が仮名文字だけで構成されている場合は、ステップＳ２２において、認識された仮名文字をその文字列の読み方とする。このとき、長音化等の発音を若干修正する場合もある。

【００９５】一方、ステップＳ２１において、画像処理文字認識部１４０によって認識された文字列が仮名文字以外の文字を含んでいる場合、ステップＳ２３において、その文字列が単語読み属性テーブル１５１に含まれているか否かを判別する。

【００９６】文字列が単語読み属性テーブル１５１に含まれている場合は、そのテーブルから読み仮名を取得し、更に発音のしかた（読み方）を生成する（ステップＳ２４）。また、単語読み属性テーブル１５１に単語の属性が記述されている場合は、属性も同時に取得する。この属性の利用方法については、後述する。

【００９７】文字列が単語読み属性テーブル１５１に含まれていない場合、ステップＳ２５において、最長一致法・分割最小法、文字読みテーブル１５２に基づく読み付与、及び読み付与規則に基づく読み付与を組み合わせて読み仮名を取得する。

【００９８】最長一致法・分割数最小法とは、単語読み属性テーブル１５１に含まれる単語を複数組み合わせることで入力文字列と同じものが構成できないか試みる方法である。例えば、入力文字列が「北品川駅前」である場合、これが単語読み属性テーブル１５１に含まれていなくても「北品川」と「駅前」とが含まれていれば、これらの組み合わせから「北品川駅前」が構成できることから、結果として「きたしながわえきまえ」という読み方が取得できる。構成方法が複数通りある場合は、より長い単語が含まれる方を優先する（最長一致法）か、より少ない単語で構成できる方を優先する（分割数最小法）かして構成方法を選択する。

【００９９】また、文字読みテーブル１５２に基づく読み付与とは、文字列を文字毎に分割し、分割した文字毎に文字読みテーブル１５２から読み仮名を取得する方法である。漢字の場合、１つの漢字には複数の読み仮名が付与できるため、文字列全体としての読み仮名は、各漢字の読み仮名の組み合わせになる。そのため、例えば、「音読みと訓読とは混在しにくい」等の規則を用いて組み合わせの数を減らす方法である。

【０１００】続いて、ステップＳ２６において、上述の各方法で取得したそれぞれの読み仮名の候補に対してスコア又は信頼度を計算し、高いものを選択する。それにより、入力された文字列に読み仮名を付与できる。得られた読み仮名から発音のしかた（読み方）を生成する。

【０１０１】ステップＳ２２、ステップＳ２４、ステップＳ２６のそれぞれの工程を経たのち、最終的に、ステップＳ２７において、読み仮名に対する発音のしかた（読み方）を長音化や連濁化等の規則に基づいて修正する。

【０１０２】ここで、単語読み属性テーブル１５１について詳細に説明する。音声認識用辞書１２１に単語を新規登録しただけでは、言語モデル１２３に記録された単語間の接続規則を適用することはできない。例えば、「北品川」を音声認識用辞書１２１に追加登録したとしても、それだけでは「北品川」に関する文法や「北品川」と他の単語との連鎖確率等は、生成されない。したがって、新規登録語に言語モデルの接続規則を反映させる方法は、理想的には、文法を追加したり、テキストデータから連鎖確率を計算し直したりして、言語モデルを構成し直すことであるが、以下に示す簡易的な方法によって新規登録後に言語モデルを適用することができる。

【０１０３】まず、言語モデルに含まれていない単語に＜未知語＞というクラス名を付ける。言語モデルには＜未知語＞と他の単語との連鎖確率を記述しておく。新規登録語は、＜未知語＞とみなし、この新規登録語と他の単語との連鎖確率は、＜未知語＞と他の単語との連鎖確率から計算する。

【０１０４】クラスとは、単語をある基準や属性にしたがって分類したものである。例えば、意味にしたがって

分類し、それぞれを＜地名＞、＜姓＞、＜国名＞と命名したり、品詞にしたがって分類し、それぞれを＜名詞＞、＜動詞＞、＜形容詞＞と命名したりする。

【０１０５】言語モデルには、単語間の連鎖確率を記述するかわりにクラス間の連鎖確率やクラスと単語との連鎖確率を記述する。単語間の連鎖確率を求めるときは、単語がどのクラスに属すかを調べ、次に対応するクラスについての連鎖確率を求め、そこから単語間の連鎖確率を計算する。

【０１０６】新規登録語についても、どのクラスに属する単語であるかを登録時に推定することでクラスモデルが適用できる。

【０１０７】上述のようにすると未知語用モデルでは、新規登録語には、全て同一の値の連鎖確率が付される。それに対してクラスモデルでは、どのクラスに属するかによって異なる値になる。そのため一般的には、新規登録語についての言語的スコアは、クラスモデルを用いた方がより適切なスコアとなり、結果的に適切に認識される。

【０１０８】したがって、音声認識による単語登録において、従来、困難であったクラス名称が、容易に入力できる。すなわち、文字認識で得られた文字列（単語）が単語読み属性テーブル１５１に含まれている場合、このテーブルの属性欄からクラス名称を取得できる。なお、図５に示す例では、属性欄に属性を１つしか記述してないが、これを「＜地名＞、＜固有名詞＞、＜駅名＞」のように複数記述することもできる。この場合、例えば、＜地名＞というクラスが存在する場合は、＜地名＞、＜固有名詞＞、＜駅名＞の中から、クラス名称と一致する分類名、すなわち＜地名＞を採用する。

【０１０９】文字認識では、一文字ずつ認識するよりも、文字の連鎖に関する情報を含めて認識する方が精度が向上する場合がある。そこで、認識用辞書の「単語シンボル」欄や、単語読み属性テーブル１５１の「単語」欄等を文字の連鎖に関する情報として使用することによって、文字認識の精度を更に向上できる。

【０１１０】以上の説明では、取得画像における所定パターンの認識として文字認識の場合に関して説明したが、上述したように文字（文字列）のほか、物体の形状、輪郭、柄、物体そのものの画像を認識し対応する文字（文字列）を抽出し、抽出された文字から推定される複数通りの読み仮名を付与し、付与された複数通りの読み仮名のそれぞれに相当する音声波形を生成することもできる。この場合は、図１に示した基本的な構成に加えて、必要な構成が必要に応じて追加される。

【０１１１】このように、所定パターンとして文字列以外にも種々のケースに対応して発音のしかたをマスターできるようにすることにより、ロボット装置が外部から情報を得て学習していく様子を表現でき、エンターテインメント性が向上できる。

【０１１２】ところで、本実施の形態として示すロボット装置１は、内部状態に応じて自律的に行動できるロボット装置である。ロボット装置１における制御プログラムのソフトウェア構成は、図１０に示すようになる。この制御プログラムは、上述したように、予めフラッシュＲＯＭ１２に格納されており、ロボット装置１の電源投入初期時において読み出される。

【０１１３】図１０において、デバイス・ドライバ・レイヤ３０は、制御プログラムの最下位層に位置し、複数のデバイス・ドライバからなるデバイス・ドライバ・セット３１から構成されている。この場合、各デバイス・ドライバは、ＣＣＤカメラ２０（図２）やタイマ等の通常のコンピュータで用いられるハードウェアに直接アクセスすることを許されたオブジェクトであり、対応するハードウェアからの割り込みを受けて処理を行う。

【０１１４】また、ロボティック・サーバ・オブジェクト３２は、デバイス・ドライバ・レイヤ３０の最下位層に位置し、例えば上述の各種センサやアクチュエータ２５$_1$～２５$_n$等のハードウェアにアクセスするためのインターフェイスを提供するソフトウェア群でなるバーチャル・ロボット３３と、電源の切換え等を管理するソフトウェア群でなるパワーマネージャ３４と、他の種々のデバイス・ドライバを管理するソフトウェア群でなるデバイス・ドライバ・マネージャ３５と、ロボット装置１の機構を管理するソフトウェア群でなるデザインド・ロボット３６とから構成されている。

【０１１５】マネージャ・オブジェクト３７は、オブジェクト・マネージャ３８及びサービス・マネージャ３９から構成されている。オブジェクト・マネージャ３８は、ロボティック・サーバ・オブジェクト３２、ミドル・ウェア・レイヤ４０、及びアプリケーション・レイヤ４１に含まれる各ソフトウェア群の起動や終了を管理するソフトウェア群であり、サービス・マネージャ３９は、メモリカード２８（図２）に格納されたコネクションファイルに記述されている各オブジェクト間の接続情報に基づいて各オブジェクトの接続を管理するソフトウェア群である。

【０１１６】ミドル・ウェア・レイヤ４０は、ロボティック・サーバ・オブジェクト３２の上位層に位置し、画像処理や音声処理等のこのロボット装置１の基本的な機能を提供するソフトウェア群から構成されている。また、アプリケーション・レイヤ４１は、ミドル・ウェア・レイヤ４０の上位層に位置し、当該ミドル・ウェア・レイヤ４０を構成する各ソフトウェア群によって処理された処理結果に基づいてロボット装置１の行動を決定するためのソフトウェア群から構成されている。

【０１１７】なお、ミドル・ウェア・レイヤ４０及びアプリケーション・レイヤ４１の具体なソフトウェア構成をそれぞれ図１１に示す。

【０１１８】ミドル・ウェア・レイヤ４０は、図１１に

示すように、騒音検出用、温度検出用、明るさ検出用、音階認識用、距離検出用、姿勢検出用、タッチセンサ用、動き検出用及び色認識用の各信号処理モジュール５０～５８並びに入力セマンティクスコンバータモジュール５９等を有する認識系６０と、出力セマンティクスコンバータモジュール６８並びに姿勢管理用、トラッキング用、モーション再生用、歩行用、転倒復帰用、ＬＥＤ点灯用及び音再生用の各信号処理モジュール６１～６７等を有する出力系６９とから構成されている。

【０１１９】認識系６０の各信号処理モジュール５０～５８は、ロボティック・サーバ・オブジェクト３２のバーチャル・ロボット３３によりＤＲＡＭ１１（図2）から読み出される各センサデータや画像データ及び音声データのうちの対応するデータを取り込み、当該データに基づいて所定の処理を施して、処理結果を入力セマンティクスコンバータモジュール５９に与える。ここで、例えば、バーチャル・ロボット３３は、所定の通信規約によって、信号の授受或いは変換をする部分として構成されている。

【０１２０】入力セマンティクスコンバータモジュール５９は、これら各信号処理モジュール５０～５８から与えられる処理結果に基づいて、「うるさい」、「暑い」、「明るい」、「ボールを検出した」、「転倒を検出した」、「撫でられた」、「叩かれた」、「ドミソの音階が聞こえた」、「動く物体を検出した」又は「障害物を検出した」等の自己及び周囲の状況や、使用者からの指令及び働きかけを認識し、認識結果をアプリケーション・レイヤ４１に出力する。

【０１２１】アプリケーション・レイヤ４１は、図１２に示すように、行動モデルライブラリ７０、行動切換えモジュール７１、学習モジュール７２、感情モデル７３及び本能モデル７４の５つのモジュールから構成されている。

【０１２２】行動モデルライブラリ７０には、図１３に示すように、「バッテリ残量が少なくなった場合」、「転倒復帰する」、「障害物を回避する場合」、「感情を表現する場合」、「ボールを検出した場合」等の予め選択されたいくつかの条件項目にそれぞれ対応させて、それぞれ独立した行動モデルが設けられている。

【０１２３】そして、これら行動モデルは、それぞれ入力セマンティクスコンバータモジュール５９から認識結果が与えられたときや、最後の認識結果が与えられてから一定時間が経過したとき等に、必要に応じて後述のように感情モデル７３に保持されている対応する情動のパラメータ値や、本能モデル７４に保持されている対応する欲求のパラメータ値を参照しながら続く行動をそれぞれ決定し、決定結果を行動切換えモジュール７１に出力する。

【０１２４】なお、この実施の形態の場合、各行動モデルは、次の行動を決定する手法として、図１４に示すような１つのノード（状態）$NODE_0$～$NODE_n$から他のどのノード$NODE_0$～$NODE_n$に遷移するかを各ノード$NODE_0$～$NODE_n$に間を接続するアークARC_1～ARC_{n1}に対してそれぞれ設定された遷移確率P_1～P_nに基づいて確率的に決定する有限確率オートマトンと呼ばれるアルゴリズムを用いる。

【０１２５】具体的に、各行動モデルは、それぞれ自己の行動モデルを形成するノード$NODE_0$～$NODE_n$にそれぞれ対応させて、これらノード$NODE_0$～$NODE_n$毎に図１５に示すような状態遷移表８０を有している。

【０１２６】この状態遷移表８０では、そのノード$NODE_0$～$NODE_n$において遷移条件とする入力イベント（認識結果）が「入力イベント名」の行に優先順に列記され、その遷移条件についての更なる条件が「データ名」及び「データ範囲」の行における対応する列に記述されている。

【０１２７】したがって、図１５の状態遷移表８０で表されるノード$NODE_{100}$では、「ボールを検出（ＢＡＬＬ）」という認識結果が与えられた場合に、当該認識結果とともに与えられるそのボールの「大きさ（ＳＩＺＥ）」が「0から1000」の範囲であることや、「障害物を検出（ＯＢＳＴＡＣＬＥ）」という認識結果が与えられた場合に、当該認識結果とともに与えられるその障害物までの「距離（ＤＩＳＴＡＮＣＥ）」が「0から100」の範囲であることが他のノードに遷移するための条件となっている。

【０１２８】また、このノード$NODE_{100}$では、認識結果の入力がない場合においても、行動モデルが周期的に参照する感情モデル７３及び本能モデル７４にそれぞれ保持された各情動及び各欲求のパラメータ値のうち、感情モデル７３に保持された「喜び（Joy）」、「驚き（Surprise）」若しくは「悲しみ（Sadness）」の何れかのパラメータ値が「50から100」の範囲であるときには他のノードに遷移することができるようになっている。

【０１２９】また、状態遷移表８０では、「他のノードへの遷移確率」の欄における「遷移先ノード」の列にそのノード$NODE_0$～$NODE_n$から遷移できるノード名が列記されているとともに、「入力イベント名」、「データ名」及び「データの範囲」の行に記述された全ての条件が揃ったときに遷移できるほかの各ノード$NODE_0$～$NODE_n$への遷移確率が「他のノードへの遷移確率」の欄内の対応する箇所にそれぞれ記述され、そのノード$NODE_0$～$NODE_n$に遷移する際に出力すべき行動が「他のノードへの遷移確率」の欄における「出力行動」の行に記述されている。なお、「他のノードへの遷移確率」の欄における各行の確率の和は１００［％］となっている。

【０１３０】したがって、図１５の状態遷移表８０で表

されるノードNODE$_{100}$　では、例えば「ボールを検出（ＢＡＬＬ）」し、そのボールの「ＳＩＺＥ（大きさ）」が「0から1000」の範囲であるという認識結果が与えられた場合には、「30［%］」の確率で「ノードＮＯＤＥ$_{120}$　（node 120）」に遷移でき、そのとき「ＡＣＴＩＯＮ１」の行動が出力されることとなる。

【０１３１】各行動モデルは、それぞれこのような状態遷移表８０として記述されたノードＮＯＤＥ$_0$ ～ ＮＯＤＥ$_n$が幾つも繋がるようにして構成されており、入力セマンティクスコンバータモジュール５９から認識結果が与えられたとき等に、対応するノードＮＯＤＥ$_0$～ＮＯＤＥ$_n$の状態遷移表を利用して確率的に次の行動を決定し、決定結果を行動切換えモジュール７１に出力するようになされている。

【０１３２】図１２に示す行動切換えモジュール７１は、行動モデルライブラリ７０の各行動モデルからそれぞれ出力される行動のうち、予め定められた優先順位の高い行動モデルから出力された行動を選択し、当該行動を実行すべき旨のコマンド（以下、これを行動コマンドという。）をミドル・ウェア・レイヤ４０の出力セマンティクスコンバータモジュール６８に送出する。なお、この実施の形態においては、図１３において下側に表記された行動モデルほど優先順位が高く設定されている。

【０１３３】また、行動切換えモジュール７１は、行動完了後に出力セマンティクスコンバータモジュール６８から与えられる行動完了情報に基づいて、その行動が完了したことを学習モジュール７２、感情モデル７３及び本能モデル７４に通知する。

【０１３４】一方、学習モジュール７２は、入力セマンティクスコンバータモジュール５９から与えられる認識結果のうち、「叩かれた」や「撫でられた」等、使用者からの働きかけとして受けた教示の認識結果を入力する　＊る。

【０１３５】そして、学習モジュール７２は、この認識結果及び行動切換えモジュール７１からの通知に基づいて、「叩かれた（叱られた）」ときにはその行動の発現確率を低下させ、「撫でられた（誉められた）」ときにはその行動の発現確率を上昇させるように、行動モデルライブラリ７０における対応する行動モデルの対応する遷移確率を変更する。

【０１３６】他方、感情モデル７３は、「喜び（Joy）」、「悲しみ（Sadness）」、「怒り（Anger）」、「驚き（Surprise）」、「嫌悪（Disgust）」及び「恐れ（Fear）」の合計６つの情動について、各情動毎にその情動の強さを表すパラメータを保持している。そして、感情モデル７３は、これら各情動のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール５９から与えられる「叩かれた」及び「撫でられた」等の特定の認識結果と、経過時間及び行動切換えモジュール７１からの通知と等に基づいて周期的に更新する。

【０１３７】具体的には、感情モデル７３は、入力セマンティクスコンバータモジュール５９から与えられる認識結果と、そのときのロボット装置１の行動と、前回更新してからの経過時間と等に基づいて所定の演算式により算出されるそのときのその情動の変動量を△Ｅ［ｔ］、現在のその情動のパラメータ値をＥ［ｔ］、その情動の感度を表す係数をk_eとして、（１）式によって次の周期におけるその情動のパラメータ値Ｅ［ｔ＋１］を算出し、これを現在のその情動のパラメータ値Ｅ［ｔ］と置き換えるようにしてその情動のパラメータ値を更新する。また、感情モデル７３は、これと同様にして全ての情動のパラメータ値を更新する。

【０１３８】

【数１】

$$E[t+1] = E[t] + ke \times \Delta E[t] \quad \cdots (1)$$

【０１３９】なお、各認識結果や出力セマンティクスコンバータモジュール６８からの通知が各情動のパラメータ値の変動量△Ｅ［ｔ］にどの程度の影響を与えるかは予め決められており、例えば「叩かれた」といった認識結果は「怒り」の情動のパラメータ値の変動量△Ｅ［ｔ］に大きな影響を与え、「撫でられた」といった認識結果は「喜び」の情動のパラメータ値の変動量△Ｅ［ｔ］に大きな影響を与えるようになっている。

【０１４０】ここで、出力セマンティクスコンバータモジュール６８からの通知とは、いわゆる行動のフィードバック情報（行動完了情報）であり、行動の出現結果の情報であり、感情モデル７３は、このような情報によっても感情を変化させる。これは、例えば、「吠える」といった行動により怒りの感情レベルが下がるといったようなことである。なお、出力セマンティクスコンバータモジュール６８からの通知は、上述した学習モジュール７２にも入力されており、学習モジュール７２は、その通知に基づいて行動モデルの対応する遷移確率を変更する。

【０１４１】なお、行動結果のフィードバックは、行動切換えモジュレータ７１の出力（感情が付加された行動）によりなされるものであってもよい。

【０１４２】一方、本能モデル７４は、「運動欲（exercise）」、「愛情欲（affection）」、「食欲（appetite）」及び「好奇心（curiosity）」の互いに独立した４つの欲求について、これら欲求毎にその欲求の強さを表すパラメータを保持している。そして、本能モデル７４は、これらの欲求のパラメータ値を、それぞれ入力セマンティクスコンバータモジュール５９から与えられる認識結果や、経過時間及び行動切換えモジュール７１からの通知等に基づいて周期的に更新する。

【０１４３】具体的には、本能モデル７４は、「運動

欲」、「愛情欲」及び「好奇心」については、認識結果、経過時間及び出力セマンティクスコンバータモジュール６８からの通知等に基づいて所定の演算式により算出されるそのときのその欲求の変動量をΔＩ［ｋ］、現在のその欲求のパラメータ値をＩ［ｋ］、その欲求の感度を表す係数ｋｉとして、所定周期で（２）式を用いて次の周期におけるその欲求のパラメータ値Ｉ［ｋ＋１］＊を算出し、この演算結果を現在のその欲求のパラメータ値Ｉ［ｋ］と置き換えるようにしてその欲求のパラメータ値を更新する。また、本能モデル７４は、これと同様にして「食欲」を除く各欲求のパラメータ値を更新する。

【０１４４】

【数２】

$$I[k+1] = I[k] + ki \times \Delta I[k] \quad \cdots (2)$$

【０１４５】なお、認識結果及び出力セマンティクスコンバータモジュール６８からの通知等が各欲求のパラメータ値の変動量△Ｉ［ｋ］にどの程度の影響を与えるかは予め決められており、例えば出力セマンティクスコンバータモジュール６８からの通知は、「疲れ」のパラメータ値の変動量△Ｉ［ｋ］に大きな影響を与えるようになっている。

【０１４６】なお、本実施の形態においては、各情動及び各欲求（本能）のパラメータ値がそれぞれ0から100までの範囲で変動するように規制されており、また係数ｋｅ、ｋｉの値も各情動及び各欲求毎に個別に設定されている。

【０１４７】一方、ミドル・ウェア・レイヤ４０の出力セマンティクスコンバータモジュール６８は、図１１に示すように、上述のようにしてアプリケーション・レイヤ４１の行動切換えモジュール７１から与えられる「前進」、「喜ぶ」、「鳴く」又は「トラッキング（ボールを追いかける）」といった抽象的な行動コマンドを出力系６９の対応する信号処理モジュール６１～６７に与える。

【０１４８】そしてこれら信号処理モジュール６１～６７は、行動コマンドが与えられると当該行動コマンドに基づいて、その行動をするために対応するアクチュエータ２５１～２５ｎ（図２）に与えるべきサーボ指令値や、スピーカ２４（図２）から出力する音の音声データ及び又は「目」のＬＥＤに与える駆動データを生成し、これらのデータをロボティック・サーバ・オブジェクト３２のバーチャル・ロボット３３及び信号処理回路１４（図２）を順次介して対応するアクチュエータ２５１～２５ｎ又はスピーカ２４又はＬＥＤに順次送出する。

【０１４９】このようにしてロボット装置１は、制御プログラムに基づいて、自己（内部）及び周囲（外部）の状況や、使用者からの指示及び働きかけに応じた自律的な行動ができる。したがって、上述した文字認識処理を実行するためプログラムを備えていないロボット装置に対しても、文字認識処理によって画像から抽出した文字の発音のしかたを音声認識処理によって周囲の音から認識された音声に基づいて決定する処理を実行するための制御プログラムを読み込ませることによって、図７に示した文字認識処理を実行させることができる。

【０１５０】このような制御プログラムは、ロボット装置が読取可能な形式で記録された記録媒体を介して提供される。制御プログラムを記録する記録媒体としては、磁気読取方式の記録媒体（例えば、磁気テープ、フロッピー（登録商標）ディスク、磁気カード）、光学読取方式の記録媒体（例えば、ＣＤ－ＲＯＭ、ＭＯ、ＣＤ－Ｒ、ＤＶＤ）等が考えられる。記録媒体には、半導体メモリ（いわゆるメモリカード（矩形型、正方形型等形状は問わない。）、ＩＣカード）等の記憶媒体も含まれる。また、制御プログラムは、いわゆるインターネット等を介して提供されてもよい。

【０１５１】これらの制御プログラムは、専用の読込ドライバ装置、又はパーソナルコンピュータ等を介して再生され、有線又は無線接続によってロボット装置１に伝送されて読み込まれる。また、ロボット装置は、半導体メモリ、又はＩＣカード等の小型化された記憶媒体のドライブ装置を備える場合、これら記憶媒体から制御プログラムを直接読み込むこともできる。ロボット装置１では、メモリカード２８から読み込むことができる。

【０１５２】なお、本発明は、上述した実施の形態のみに限定されるものではなく、本発明の要旨を逸脱しない範囲において種々の変更が可能であることは勿論である。本実施の形態では、４足歩行のロボット装置に関して説明したが、ロボット装置は、２足歩行であってもよく、更に、移動手段は、脚式移動方式に限定されない。

【０１５３】以下に、本発明の別の実施の形態として示す人間型ロボット装置の詳細について説明する。図１６及び図１７には、人間型ロボット装置２００を前方及び後方の各々から眺望した様子を示している。更に、図１８には、この人間型ロボット装置２００が具備する関節自由度構成を模式的に示している。

【０１５４】図１６に示すように、人間型ロボット装置２００は、２本の腕部と頭部２０１を含む上肢と、移動動作を実現する２本の脚部からなる下肢と、上肢と下肢とを連結する体幹部とで構成される。

【０１５５】頭部２０１を支持する首関節は、首関節ヨー軸２０２と、首関節ピッチ軸２０３と、首関節ロール軸２０４という３自由度を有している。

【0156】また、各腕節は、肩関節ピッチ軸208と、肩関節ロール軸209と、上腕ヨー軸210と、肘関節ピッチ軸211と、前腕ヨー軸212と、手首関節ピッチ軸213と、手首関節ロール輪214と、手部215とで構成される。手部215は、実際には、複数本の指を含む多関節・多自由度構造体である。ただし、手部215の動作は人間型ロボット装置200の姿勢制御や歩行制御に対する寄与や影響が少ないので、本明細書ではゼロ自由度と仮定する。したがって、各腕部は7自由度を有するとする。

【0157】また、体幹部は、体幹ピッチ軸205と、体幹ロール軸206と、体幹ヨー軸207という3自由度を有する。

【0158】また、下肢を構成する各々の脚部は、股関節ヨー軸216と、股関節ピッチ軸217と、股関節ロール軸218と、膝関節ピッチ軸219と、足首関節ピッチ軸220と、足首関節ロール軸221と、足部222とで構成される。本明細書中では、股関節ピッチ軸217と股関節ロール軸218の交点は、人間型ロボット装置200の股関節位置を定義する。人体の足部222は、実際には多関節・多自由度の足底を含んだ構造体であるが、人間型ロボット装置200の足底は、ゼロ自由度とする。したがって、各脚部は、6自由度で構成される。

【0159】以上を総括すれば、人間型ロボット装置200全体としては、合計で3＋7×2＋3＋6×2＝32自由度を有することになる。ただし、エンターテインメント向けの人間型ロボット装置200が必ずしも32自由度に限定される訳ではない。設計・制作上の制約条件や要求仕様等に応じて、自由度すなわち関節数を適宜増減することができることはいうまでもない。

【0160】上述したような人間型ロボット装置200がもつ各自由度は、実際にはアクチュエータを用いて実装される。外観上で余分な膨らみを排してヒトの自然体形状に近似させること、2足歩行という不安定構造体に対して姿勢制御を行うことなどの要請から、アクチュエータは小型且つ軽量であることが好ましい。

【0161】図19には、人間型ロボット装置200の制御システム構成を模式的に示している。同図に示すように、人間型ロボット装置200は、ヒトの四肢を表現した各機構ユニット230，240，250R／L，260R／Lと、各機構ユニット間の協調動作を実現するための適応制御を行う制御ユニット280とで構成される（ただし、R及びLの各々は、右及び左の各々を示す接尾辞である。以下同様）。

【0162】人間型ロボット装置200全体の動作は、制御ユニット280によって統括的に制御される。制御ユニット280は、CPU（Central Processing Unit）やメモリ等の主要回路コンポーネント（図示しない）で構成される主制御部281と、電源回路や人間型ロボット装置200の各構成要素とのデータやコマンドの授受を行うインターフェイス（何れも図示しない）などを含んだ周辺回路282とで構成される。この制御ユニット280の設置場所は、特に限定されない。図19では体幹部ユニット240に搭載されているが、頭部ユニット230に搭載してもよい。或いは、人間型ロボット装置200外に制御ユニット280を配備して、人間型ロボット装置200の機体とは有線若しくは無線で交信するようにしてもよい。

【0163】図19に示した人間型ロボット装置200内の各関節自由度は、それぞれに対応するアクチュエータによって実現される。すなわち、頭部ユニット230には、首関節ヨー軸202、首関節ピッチ203、首関節ロール軸204の各々を表現する首関節ヨー軸アクチュエータA_2、首関節ピッチ軸アクチュエータA_3、首関節ロール軸アクチュエータA_4が配設されている。

【0164】また、頭部ユニット230には、外部の状況を撮像するためのCCD（ChargeCoupled Device）カメラが設けられているほか、前方に位置する物体までの距離を測定するための距離センサ、外部音を集音するためのマイク、音声を出力するためのスピーカ、使用者からの「撫でる」や「叩く」といった物理的な働きかけにより受けた圧力を検出するためのタッチセンサ等が配設されている。

【0165】また、体幹部ユニット240には、体幹ピッチ軸205、体幹ロール軸206、体幹ヨー軸207の各々を表現する体幹ピッチ軸アクチュエータA_5、体幹ロール軸アクチュエータA_6、体幹ヨー軸アクチュエータA_7が配設されている。また、体幹部ユニット240には、この人間型ロボット装置200の起動電源となるバッテリを備えている。このバッテリは、充放電可能な電池によって構成されている。

【0166】また、腕部ユニット250R／Lは、上腕ユニット251R／Lと、肘関節ユニット252R／Lと、前腕ユニット253R／Lに細分化されるが、肩関節ピッチ軸8、肩関節ロール軸209、上腕ヨー軸210、肘関節ピッチ軸211、前腕ヨー軸212、手首関節ピッチ軸213、手首関節ロール軸214の各々表現する肩関節ピッチ軸アクチュエータA_8、肩関節ロール軸アクチュエータA_9、上腕ヨー軸アクチュエータA_{10}、肘関節ピッチ軸アクチュエータA_{11}、肘関節ロール軸アクチュエータA_{12}、手首関節ピッチ軸アクチュエータA_{13}、手首関節ロール軸アクチュエータA_{14}が配備されている。

【0167】また、脚部ユニット260R／Lは、大腿部ユニット261R／Lと、膝ユニット262R／Lと、脛部ユニット263R／Lに細分化されるが、股関節ヨー軸216、股関節ピッチ軸217、股関節ロール軸218、膝関節ピッチ軸219、足首関節ピッチ軸220、足首関節ロール軸221の各々を表現する股関節

ヨー軸アクチュエータA_{16}、股関節ピッチ軸アクチュエータA_{17}、股関節ロール軸アクチュエータA_{18}、膝関節ピッチ軸アクチュエータA_{19}、足首関節ピッチ軸アクチュエータA_{20}、足首関節ロール軸アクチュエータA_{21}が配備されている。各関節に用いられるアクチュエータA_2，A_3・・・は、より好ましくは、ギア直結型で且つサーボ制御系をワンチップ化してモータ・ユニット内に搭載したタイプの小型ＡＣサーボ・アクチュエータで構成することができる。

【０１６８】頭部ユニット２３０、体幹部ユニット２４０、腕部ユニット２５０、各脚部ユニット２６０などの各機構ユニット毎に、アクチュエータ駆動制御部の副制御部２３５，２４５，２５５Ｒ／Ｌ，２６５Ｒ／Ｌが配備されている。更に、各脚部２６０Ｒ，Ｌの足底が着床したか否かを検出する接地確認センサ２９１及び２９２を装着するとともに、体幹部ユニット２４０内には、姿勢を計測する姿勢センサ２９３を装備している。

【０１６９】接地確認センサ２９１及び２９２は、例えば足底に設置された近接センサ又はマイクロ・スイッチなどで構成される。また、姿勢センサ２９３は、例えば、加速度センサとジャイロ・センサの組み合わせによって構成される。

【０１７０】接地確認センサ２９１及び２９２の出力によって、歩行・走行などの動作期間中において、左右の各脚部が現在立脚又は遊脚何れの状態であるかを判別することができる。また、姿勢センサ２９３の出力により、体幹部分の傾きや姿勢を検出することができる。

【０１７１】主制御部２８１は、各センサ２９１～２９３の出力に応答して制御目標をダイナミックに補正することができる。より具体的には、副制御部２３５，２４５，２５５Ｒ／Ｌ，２６５Ｒ／Ｌの各々に対して適応的な制御を行い、人間型ロボット装置２００の上肢、体幹、及び下肢が協調して駆動する全身運動パターンを実現できる。

【０１７２】人間型ロボット装置２００の機体上での全身運動は、足部運動、ＺＭＰ（ZeroMoment Point）軌道、体幹運動、上肢運動、腰部高さなどを設定するとともに、これらの設定内容にしたがった動作を指示するコマンドを各副制御部２３５，２４５，２５５Ｒ／Ｌ，２６５Ｒ／Ｌに転送する。そして、各々の副制御部２３５，２４５，・・・等では、主制御部２８１からの受信コマンドを解釈して、各アクチュエータA_2，A_3・・・等に対して駆動制御信号を出力する。ここでいう「ＺＭＰ」とは、歩行中の床反力によるモーメントがゼロとなる床面上の点のことであり、また、「ＺＭＰ軌道」とは、例えば人間型ロボット装置２００の歩行動作期間中にＺＭＰが動く軌跡を意味する。

【０１７３】歩行時には、重力と歩行運動に伴って生じる加速度によって、歩行系から路面には重力と慣性力、並びにこれらのモーメントが作用する。いわゆる「ダランベールの原理」によると、それらは路面から歩行系への反作用としての床反力、床反力モーメントとバランスする。力学的推論の帰結として、足底接地点と路面の形成する支持多角形の辺上或いはその内側にピッチ及びロール軸モーメントがゼロとなる点、すなわち「ＺＭＰ（Zero Moment Point）」が存在する。

【０１７４】脚式移動ロボットの姿勢安定制御や歩行時の転倒防止に関する提案の多くは、このＺＭＰを歩行の安定度判別の規範として用いたものである。ＺＭＰ規範に基づく２足歩行パターン生成は、足底着地点を予め設定することができ、路面形状に応じた足先の運動学的拘束条件を考慮しやすいなどの利点がある。また、ＺＭＰを安定度判別規範とすることは、力ではなく軌道を運動制御上の目標値として扱うことを意味するので、技術的に実現可能性が高まる。なお、ＺＭＰの概念並びにＺＭＰを歩行ロボットの安定度判別規範に適用する点については、Miomir Vukobratovic著“LEGGED LOCOMOTION ROBOTS”（加藤一郎外著『歩行ロボットと人工の足』（日刊工業新聞社））に記載されている。

【０１７５】一般には、４足歩行よりもヒューマノイドのような２足歩行のロボットの方が、重心位置が高く、且つ、歩行時のＺＭＰ安定領域が狭い。したがって、このような路面状態の変化に伴う姿勢変動の問題は、２足歩行ロボットにおいてとりわけ重要となる。

【０１７６】以上のように、人間型ロボット装置２００は、各々の副制御部２３５，２４５，・・・等が、主制御部２８１からの受信コマンドを解釈して、各アクチュエータA_2，A_3・・・に対して駆動制御信号を出力し、各ユニットの駆動を制御している。これにより、人間型ロボット装置２００は、目標の姿勢に安定して遷移し、安定した姿勢で歩行できる。

【０１７７】また、人間型ロボット装置２００における制御ユニット２８０では、上述したような姿勢制御のほかに、加速度センサ、タッチセンサ、接地確認センサ等の各種センサ、及びＣＣＤカメラからの画像情報、マイクからの音声情報等を統括して処理している。制御ユニット２８０では、図示しないが加速度センサ、ジャイロ・センサ、タッチセンサ、距離センサ、マイク、スピーカなどの各種センサ、各アクチュエータ、ＣＣＤカメラ及びバッテリが各々対応するハブを介して主制御部２８１と接続されている。

【０１７８】主制御部２８１は、上述の各センサから供給されるセンサデータや画像データ及び音声データを順次取り込み、これらをそれぞれ内部インターフェイスを介してＤＲＡＭ内の所定位置に順次格納する。また、主制御部２８１は、バッテリから供給されるバッテリ残量を表すバッテリ残量データを順次取り込み、これをＤＲＡＭ内の所定位置に格納する。ＤＲＡＭに格納された各センサデータ、画像データ、音声データ及びバッテリ残量データは、主制御部２８１がこの人間型ロボット装置

２００の動作制御を行う際に利用される。

【０１７９】主制御部２８１は、人間型ロボット装置２００の電源が投入された初期時、制御プログラムを読み出し、これをＤＲＡＭに格納する。また、主制御部２８１は、上述のように主制御部２８１よりＤＲＡＭに順次格納される各センサデータ、画像データ、音声データ及びバッテリ残量データに基づいて自己及び周囲の状況や、使用者からの指示及び働きかけの有無などを判断する。更に、主制御部２８１は、この判断結果及びＤＲＡＭに格納した制御プログラムに基づいて自己の状況に応じて行動を決定するとともに、当該決定結果に基づいて必要なアクチュエータを駆動させることにより人間型ロボット装置２００に、いわゆる「身振り」、「手振り」といった行動をとらせる。

【０１８０】したがって、人間型ロボット装置２００は、制御プログラムに基づいて自己及び周囲の状況を判断し、使用者からの指示及び働きかけに応じて自律的に行動できる。また、人間型ロボット装置２００は、ＣＣＤカメラにおいて撮像された画像から抽出した文字の発音のしかた（読み方）を、抽出された文字から推定される読み方と集音マイクにおいて集音された音声とをマッチングして決定する。したがって、人間型ロボット装置２００の音声認識の精度が向上し、新規単語が音声認識用辞書に登録できる。

【０１８１】

【発明の効果】以上詳細に説明したように、本発明に係るロボット装置は、単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、被写体を撮像する撮像手段と、撮像手段において撮像された画像から所定パターンの画像を抽出する画像認識手段と、周囲の音を取得する集音手段と、集音手段において取得された音から音声を認識する音声認識手段と、画像認識手段において抽出された所定パターンの画像から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する記憶制御手段とを備える。

【０１８２】本発明に係るロボット装置は、撮像手段において撮像された画像から抽出された所定パターンの画像から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与した複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成し、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較して最も近い音声波形を抽出した文字の発音のしかたであるとして決定する。

【０１８３】したがって、本発明に係るロボット装置によれば、特に、弱い音素（例えば、語頭の／ｓ／等）を含む発音の誤認識、周囲の雑音の影響による入力音素の変化、音声区間検出の失敗等による悪影響が抑止され、新規単語を登録する際の認識精度が向上できる。これにより、正確な発音のしかたが音声認識用辞書に記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

【０１８４】また、本発明に係るロボット装置は、単語とこの単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備え、記憶制御手段が新規に記憶する文字と該文字の発音のしかたとともに単語属性を対応させて音声認識用辞書に記憶する。

【０１８５】したがって、本発明に係るロボット装置によれば、入力した音声及び出力する音声に文法規則、対話規則等を適用する上で必要となる単語属性情報をユーザが入力する必要がなくなり利便性が向上するとともに、ユーザが属性情報を知らない場合に属性情報が入力できなかったという不都合が改善される。

【０１８６】また、本発明に係る文字認識装置は、単語と該単語の発音のしかたとの対応関係が音声認識用辞書として記憶された音声認識用記憶手段と、単語と該単語の表音文字との対応関係が単語表音テーブルとして記憶された単語表音記憶手段と、被写体を撮像する撮像手段と、撮像手段において撮像された画像から所定パターンの画像を抽出する画像認識手段と、周囲の音を取得する集音手段と、集音手段において取得された音から音声を認識する音声認識手段と、画像認識手段において抽出された文字から推定される複数通りの表音文字を単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成手段と、発音情報生成手段において生成された各音声波形と音声認識手段において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして音声認識用辞書に新規に記憶する記憶制御手段とを備える。

【０１８７】したがって、本発明に係る文字認識装置によれば、特に、弱い音素（例えば、語頭の／ｓ／等）を含む発音の誤認識、周囲の雑音の影響による入力音素の変化、音声区間検出の失敗等による悪影響が抑止され、新規単語を登録する際の認識精度が向上できる。これにより、正確な発音のしかたが音声認識用辞書に記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

【０１８８】また、本発明に係る文字認識装置は、単語とこの単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備

え、記憶制御手段が新規に記憶する文字と該文字の発音のしかたとともに単語属性を対応させて音声認識用辞書に記憶する。

【０１８９】したがって、本発明に係る文字認識装置によれば、入力した音声及び出力する音声に文法規則、対話規則等を適用する上で必要となる単語属性情報をユーザが入力する必要がなくなり利便性が向上するとともに、ユーザが属性情報を知らない場合は、属性情報を入力できなかったという不都合が改善される。

【０１９０】また、本発明に係る文字認識方法は、被写体を撮像する撮像工程と、撮像工程において撮像された画像から所定パターンの画像を抽出する画像認識工程と、周囲の音を取得する集音工程と、集音工程において取得された音から音声を認識する音声認識工程と、画像認識工程において抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成工程と、発音情報生成工程において生成された各音声波形と音声認識工程において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶制御工程とを備える。

【０１９１】したがって、本発明に係る文字認識方法によれば、特に、弱い音素（例えば、語頭の／ｓ／等）を含む発音の誤認識、周囲の雑音の影響による入力音素の変化、音声区間検出の失敗等による悪影響が抑止され、新規単語を登録する際の認識精度が向上できる。これにより、正確な発音のしかたが音声認識用辞書に記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

【０１９２】また、本発明に係る文字認識方法によれば、単語とこの単語の表音文字と単語属性とを含む単語情報が単語属性テーブルとして記憶された単語情報記憶手段を備え、記憶制御手段が新規に記憶する文字と該文字の発音のしかたとともに単語属性を対応させて音声認識用辞書に記憶する。

【０１９３】したがって、本発明に係る文字認識方法によれば、入力した音声及び出力する音声に文法規則、対話規則等を適用する上で必要となる単語属性情報をユーザが入力する必要がなくなり利便性が向上するとともに、ユーザが属性情報を知らない場合は、属性情報を入力できなかったという不都合が改善される。

【０１９４】更に、本発明に係る制御プログラムは、被写体を撮像する撮像処理と、撮像処理によって撮像された画像から所定パターンの画像を抽出する画像認識処理と、周囲の音を取得する集音処理と、集音処理によって取得された音から音声を認識する音声認識処理と、画像認識処理によって抽出された文字から推定される複数通りの表音文字を単語と該単語の表音文字との対応関係が記憶された単語表音テーブルに基づいて付与し、付与された複数通りの表音文字の各々に対して発音のしかたと発音に相当する音声波形とを生成する発音情報生成処理と、発音情報生成処理によって生成された各音声波形と音声認識処理において認識された音声の音声波形とを比較し、最も近い音声波形を抽出した文字の発音のしかたであるとして単語と該単語の発音のしかたとの対応関係を記憶した音声認識用辞書に新規に記憶する記憶処理とをロボット装置に実行させる。

【０１９５】したがって、本発明に係る制御プログラムによれば、ロボット装置は、特に、弱い音素（例えば、語頭の／ｓ／等）を含む発音の誤認識、周囲の雑音の影響による入力音素の変化、音声区間検出の失敗等による悪影響が抑止され、新規単語を登録する際の認識精度が向上される。これにより、正確な発音のしかたが音声認識用辞書に記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

【０１９６】また、上述の制御プログラムを記録媒体に記録して提供することによって、この記録媒体を読込可能で画像認識手段と音声認識手段とを備える音声認識装置としての機能を有する電子機器に対して、新規単語を登録する際の認識精度が向上される。これにより、正確な発音のしかたが記憶できるため、新規単語として登録された語を認識する際の認識精度が向上する。

【図面の簡単な説明】

【図１】本発明の一構成例として示すロボット装置の外観を示す外観図である。

【図２】本発明の一構成例として示すロボット装置の構成を示す構成図である。

【図３】本発明の一構成例として示すロボット装置における画像音声認識部の構成を示す構成図である。

【図４】本発明の一構成例として示すロボット装置の音声認識用辞書を説明する図である。

【図５】本発明の一構成例として示すロボット装置の単語読み属性テーブルを説明する図である。

【図６】本発明の一構成例として示すロボット装置の文字読みテーブルを説明する図である。

【図７】本発明の一構成例として示すロボット装置が新規単語を音声認識用辞書に登録する処理を説明するフローチャートである。

【図８】本発明の一構成例として示すロボット装置の新規単語用認識用辞書を説明する図である。

【図９】本発明の一構成例として示すロボット装置が認識した文字列の発音のしかた（読み方）を生成する処理を説明するフローチャートである。

【図１０】本発明の一構成例として示すロボット装置の制御プログラムのソフトウェア構成を示す構成図である。

【図１１】本発明の一構成例として示すロボット装置の制御プログラムのうち、ミドル・ウェア・レイヤの構成を示す構成図である。

【図１２】本発明の一構成例として示すロボット装置の制御プログラムのうち、アプリケーション・レイヤの構成を示す構成図である。

【図１３】本発明の一構成例として示すロボット装置の制御プログラムのうち、行動モデルライブラリの構成を示す構成図である。

【図１４】本発明の一構成例として示すロボット装置の行動を決定するためのアルゴリズムである有限確率オートマトンを説明する模式図である。

【図１５】本発明の一構成例として示すロボット装置の行動を決定するための状態遷移条件を表す図である。

【図１６】本発明の一構成例として示す人間型ロボット装置の前方からみた外観を説明する外観図である。

【図１７】本発明の一構成例として示す人間型ロボット装置の後方からみた外観を説明する外観図である。

【図１８】本発明の一構成例として示す人間型ロボット装置の自由度構成モデルを模式的に示す図である。

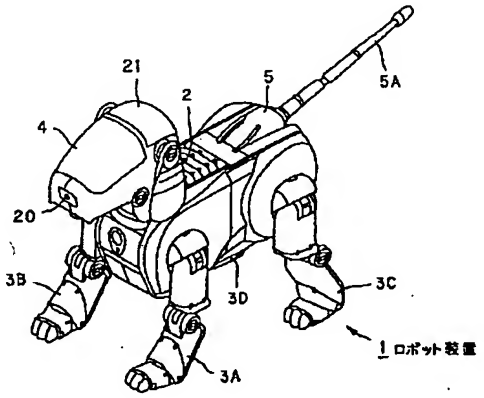
【図１９】本発明の一構成例として示す人間型ロボット装置の制御システム構成を説明する図である。

【図２０】図２０（ａ）は、「音素」を基本単位とする＊ガーベージモデルを適用した従来の音声認識方法を示す模式図であり、図２０（ｂ）は、「かな」を基本単位とするガーベージモデルを適用した従来の音声認識方法を示す模式図である。

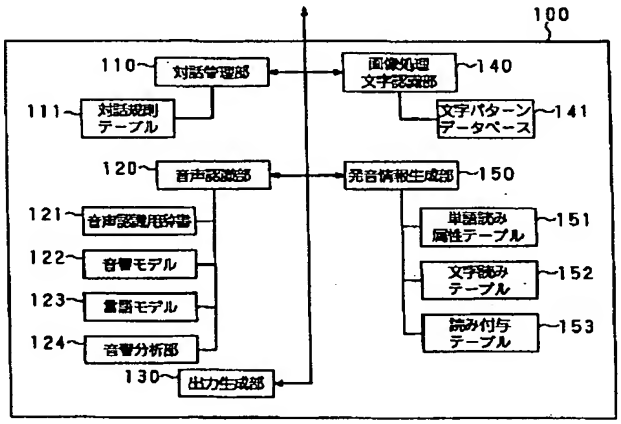
【符号の説明】

１　ロボット装置、２　胴体部ユニット、３Ａ，３Ｂ，３Ｃ，３Ｄ　脚部ユニット、４　頭部ユニット、５　尻尾部ユニット、１０　ＣＰＵ、１１　ＤＲＡＭ、１２　フラッシュＲＯＭ、１３　ＰＣカードインターフェイス回路、１４　信号処理回路、１５　内部バス、１６　コントロール部、１７　バッテリ、１８角速度センサ、１９　加速度センサ、２０　ＣＣＤカメラ、２１　タッチセンサ、２２　距離センサ、２３　マイク、２４　スピーカ、２５１～２５ｎ　アクチュエータ、２６１～２６ｎ　ポテンショメータ、２７１～２７ｎ　ハブ、２８メモリカード、１００　画像音声認識部、１１０　対話管理部、１１１対話規則テーブル、１２０　音声認識部、１２１　音声認識用辞書、１２２　音響モデル、１２３　言語モデル、１２４　音響分析部、１３０　出力生成部、１４０　画像処理文字認識部、１４１　文字パターンデータベース、１５０　発音情報生成部、１５１　単語読み属性テーブル、１５２　文字読みテーブル、１５３　読み付与テーブル、２００　人間型ロボット装置

【図１】



【図３】



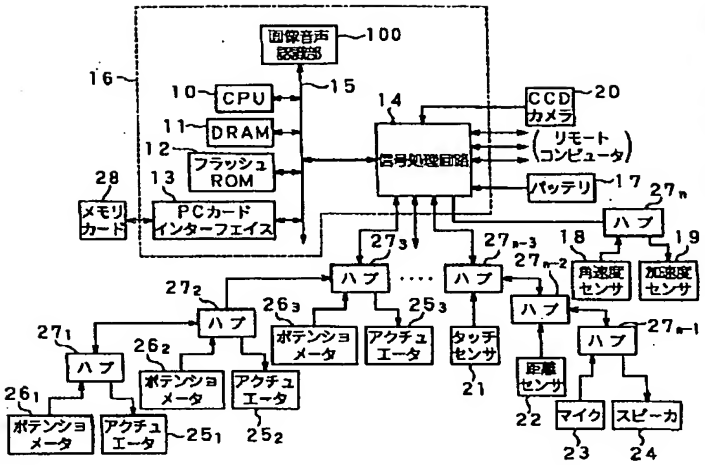
【図４】

単語シンボル	PLU列
五反田	ｇｏｔｅＮｄａ
電車	ｄｅＮｓｈａ
に	ｎｉ
行き	ｉｋｉ
たい	ｔａｉ
・	・
・	・

【図５】

単語	読み	属性
品川	しながわ	（地名）
北品川	きたしながわ	（地名）
吉物町丁	あおものよこちょう	（地名）
・	・	・
・	・	・
佐藤	さとう	（姓）
鈴木	すずき	（姓）
・	・	・
・	・	・
すべすべ猫図猫	すべすべまんじゅうがに	（動物）
大猫海豚	おおいかりなまこ	（動物）
・	・	・
・	・	・

【図２】



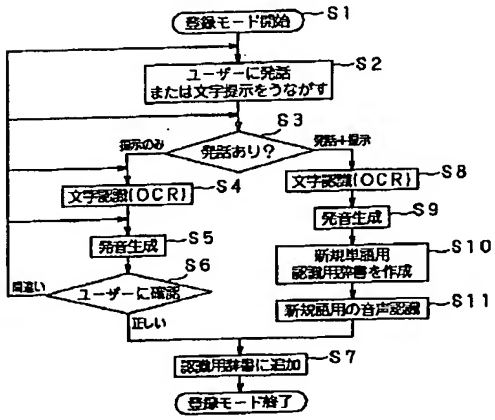
【図６】

単語	読み
北	きた,ホク
品	しな,ヒン,ホン
川	かわ,セン
比	くら,ヒ
品	おぎら,ショウ
：	：
？	はてな,クエスチョン
@	アット
：	：
A	エー,ア
B	ビー,ブ
：	：

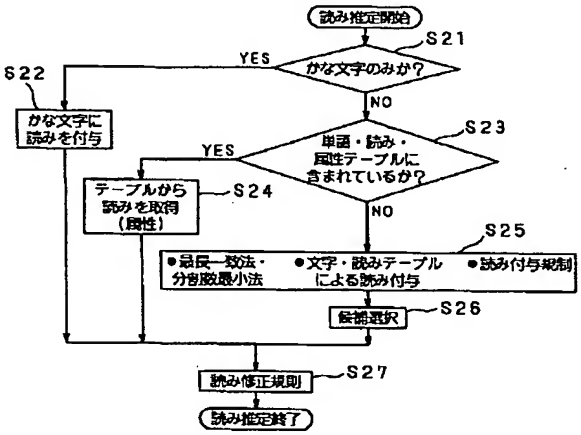
【図８】

単語シンボル	PLU列
北品川	k i t a s h i n a g a w a
北品川	h i s h o : k a w a
比品川	k u r a s k i r e g a w a

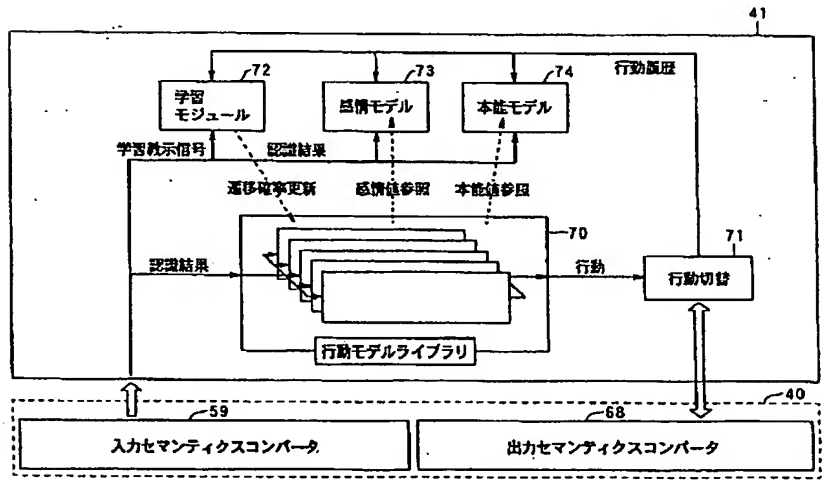
【図７】



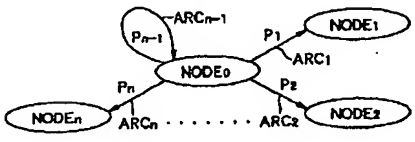
【図９】



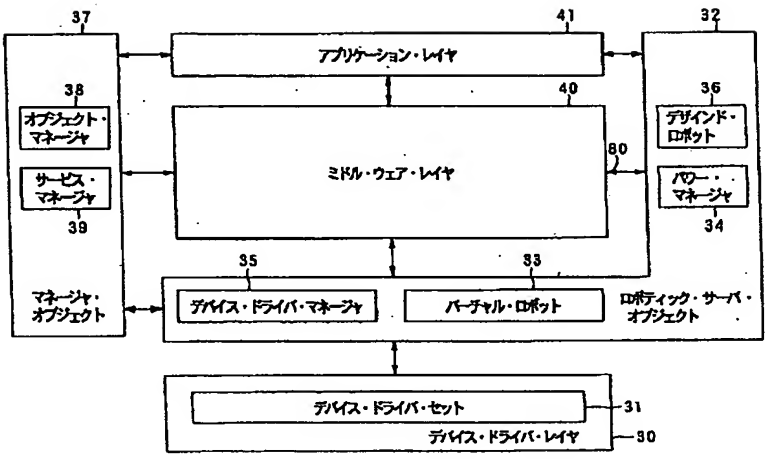
【図１２】



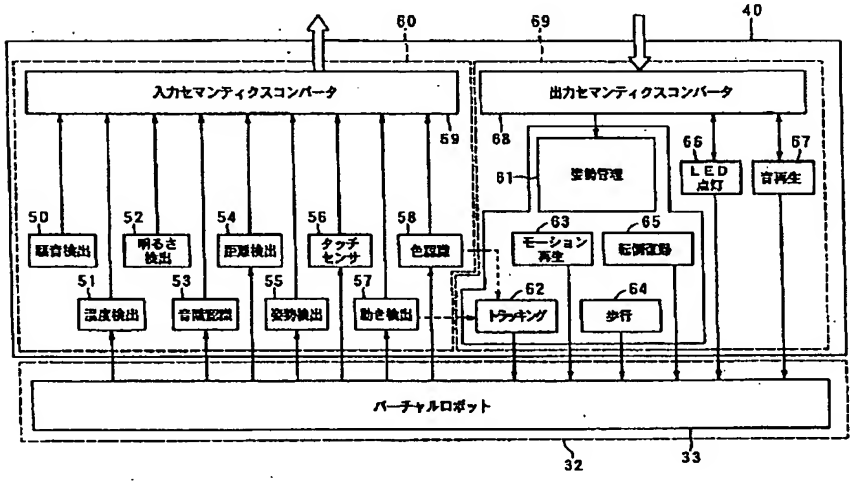
【図１４】



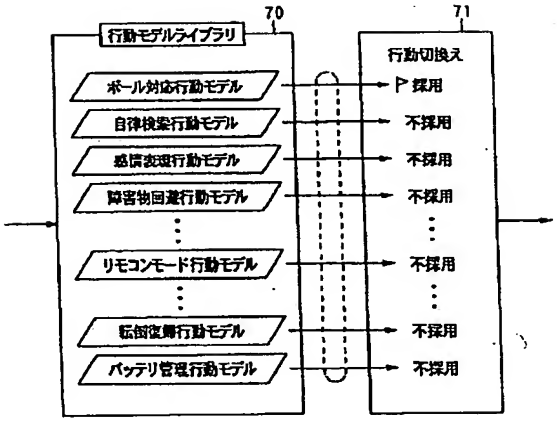
【図１０】


37
41
32
アプリケーション・レイヤ
38
オブジェクト・マネージャ
40
36
デザインド・ロボット
ミドル・ウェア・レイヤ
80
サービス・マネージャ
パワー・マネージャ
39
34
35
33
マネージャ・オブジェクト
デバイス・ドライバ・マネージャ
バーチャル・ロボット
ロボティック・サーバ・オブジェクト
デバイス・ドライバ・セット
31
デバイス・ドライバ・レイヤ
30

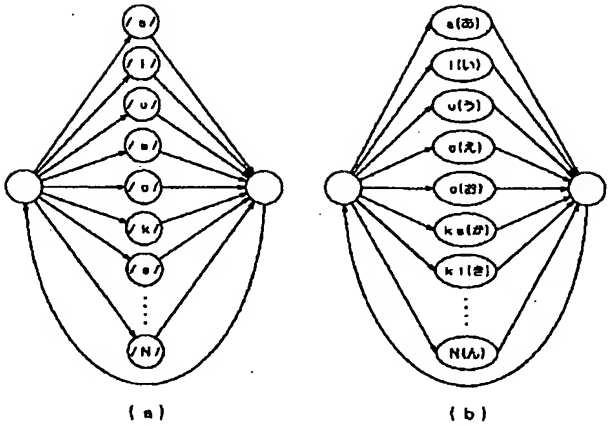
【図１１】


60
69
40
入力セマンティクスコンバータ
出力セマンティクスコンバータ
59
68
66
67
61
姿勢管理
LED点灯
音再生
50
52
54
56
58
63
65
顔音検出
明るさ検出
距離検出
タッチセンサ
色認識
モーション再生
転倒復帰
51
53
55
57
62
64
深度検出
音源認識
姿勢検出
動き検出
トラッキング
歩行
バーチャルロボット
32
33

【図１３】



【図２０】

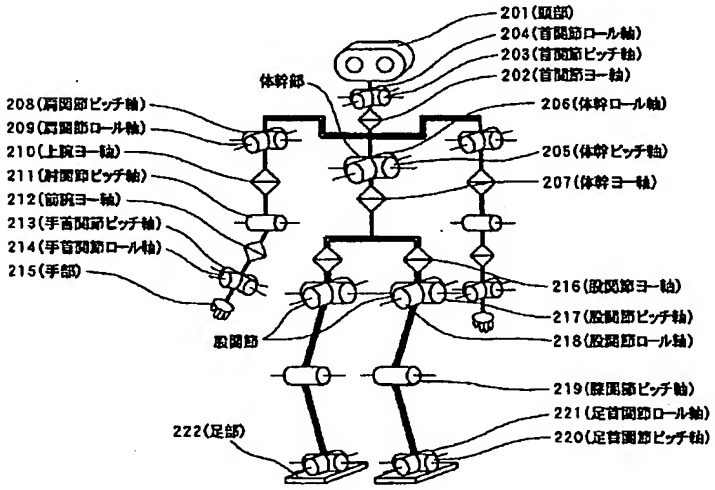


【図１５】

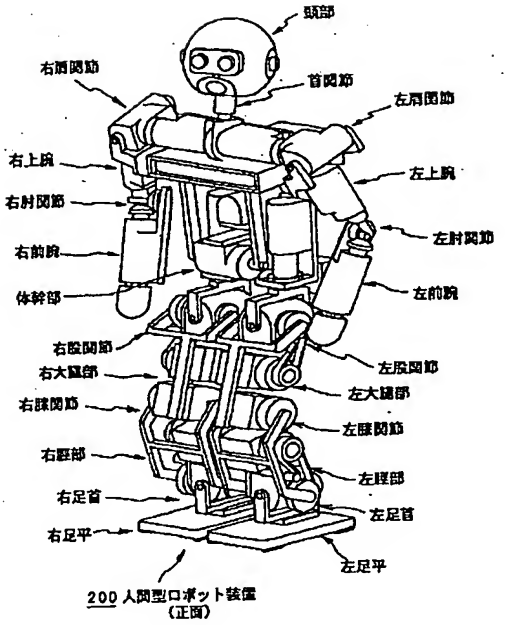
	入力イベント名	データ名	データの範囲	他のノードへの遷移確率 DI				
node 100				A	B	C	D	n
遷移先ノード				node 120	node 120	node 1000		node 600
出力行動				ACTION 1	ACTION 2	MOVE BACK		ACTION 4
1	BALL	SIZE	0.1000	30%				
2	PAT				40%			
3	HIT				20%			
4	MOTION					50%		
5	OBSTACLE	DISTANCE	0.100			100%		
6		JOY	50.100					
7		SURPRISE	50.100					
8		SADNESS	50.100					

80

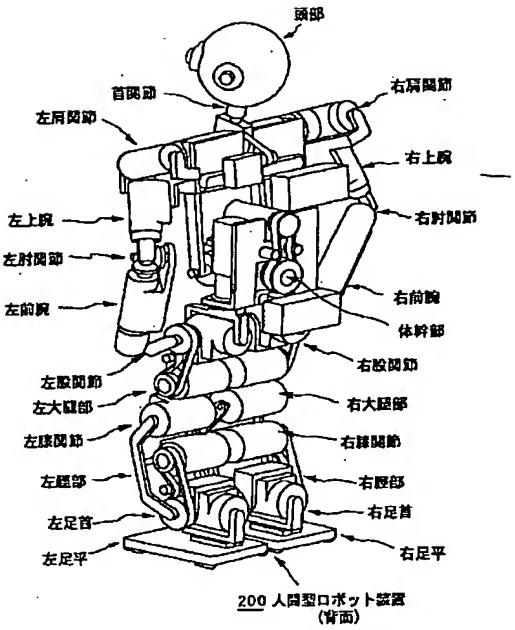
【図１８】



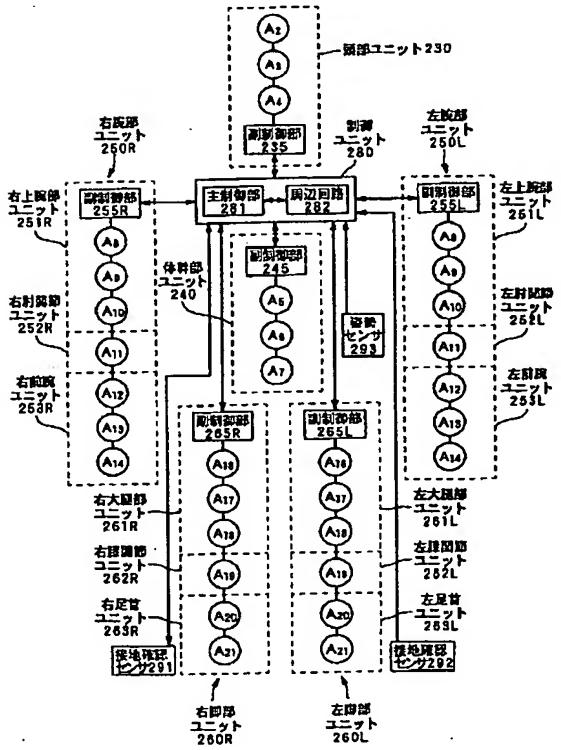
【図１６】


頭部
首関節
右肩関節
左肩関節
右上腕
左上腕
右肘関節
左肘関節
右前腕
左前腕
体幹部
右股関節
左股関節
右大腿部
左大腿部
右膝関節
左膝関節
右脛部
左脛部
右足首
左足首
右足平
左足平
200 人間型ロボット装置
(正面)

【図１７】


頭部
首関節
右肩関節
左肩関節
右上腕
左上腕
右肘関節
左肘関節
右前腕
左前腕
体幹部
右股関節
左股関節
右大腿部
左大腿部
右膝関節
左膝関節
右脛部
左脛部
右足首
左足首
右足平
左足平
200 人間型ロボット装置
(背面)

【図１９】


頭部ユニット230
制御ユニット280
主制御部281
周辺回路282
右脚部ユニット260R
左脚部ユニット260L
体幹部ユニット240
接地確認センサ291
接地確認センサ292
接地確認センサ293

フロントページの続き

(51)Int.Cl.[7]	識別記号	FI		ﾃｰﾏｺｰﾄﾞ(参考)
Ｇ１０Ｌ	15/22	Ｇ１０Ｌ	3/00	５２１Ｖ
	15/24			５７１Ｑ
				５７１Ｔ
				５３１Ｑ

(72)発明者　河本　献太
東京都品川区北品川６丁目７番35号　ソニー株式会社内
(72)発明者　佐部　浩太郎
東京都品川区北品川６丁目７番35号　ソニー株式会社内
(72)発明者　大橋　武史
東京都品川区北品川６丁目７番35号　ソニー株式会社内

Ｆターム(参考)　5B064 AA07 FA16
5D015 GG03 HH23 KK02 KK04 LL07 LL11

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-044080

(43)Date of publication of application : 14.02.2003

---

(51)Int.Cl. G10L 15/06

G06K 9/00

G10L 15/00

G10L 15/14

G10L 15/20

G10L 15/22

G10L 15/24

---

(21)Application number : 2002-130905 (71)Applicant : SONY CORP

(22)Date of filing : 02.05.2002 (72)Inventor : HIROE ATSUO

MINAMINO KATSUKI

KAWAMOTO KENTA

SABE KOTARO

OHASHI TAKESHI

---

(30)Priority

Priority number : 2001135423

Priority date : 02.05.2001

Priority country : JP

---

(54) ROBOT DEVICE, DEVICE AND METHOD FOR RECOGNIZING CHARACTER, CONTROL PROGRAM AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To register a unregistered word at a dictionary for

recognition as a new word.

SOLUTION: A plurality of characters estimated from the result of the character recognition of an image picked up by a CCD camera 20, a plurality of Kanas for the reading of characters estimated from these respective characters and reading corresponding to each of Kanas for the reading of characters are generated in a pronunciation information generating part 150 and by matching a plurality of readings provided therein and voice acquired from the user by a microphone 23, one Kana for the reading of character and one of pronunciation (reading) are specified out of a plurality of generated candidates.

LEGAL STATUS [Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

CLAIMS

[Claim(s)]

[Claim 1] A storage means for speech recognition by which correspondence relation with the method of the pronunciation of a word and this word was memorized as a dictionary for speech recognition in the robot equipment which operates autonomously according to an internal state, A word phonetic storage means by which correspondence relation with the phonogram of a word and this word was memorized as a word phonetic table, An image pick-up means to

picturize a photographic subject, and an image recognition means to extract the image of a predetermined pattern from the image picturized in the above-mentioned image pick-up means, A sound-collecting means to acquire a surrounding sound, and a speech recognition means to recognize voice from the sound acquired in the above-mentioned sound-collecting means, Two or more kinds of phonograms presumed from the above-mentioned predetermined pattern extracted in the above-mentioned image recognition means are given based on the above-mentioned word phonetic table. A pronunciation information generation means to generate the method of pronunciation, and the voice wave equivalent to pronunciation to each of two or more kinds of phonograms by which grant was carried out [above-mentioned], The voice wave of the voice recognized in each voice wave and the above-mentioned speech recognition means which were generated in the above-mentioned pronunciation information generation means is compared. Robot equipment characterized by having a storage control means to memorize newly in the above-mentioned dictionary for speech recognition noting that it is the method of the pronunciation corresponding to the pattern recognition result from which the nearest voice wave was extracted in the above-mentioned image recognition means.

[Claim 2] The image of the above-mentioned predetermined pattern is robot equipment according to claim 1 characterized by being the character string

which consists of an alphabetic character and/or two or more alphabetic characters.

[Claim 3] Robot equipment according to claim 2 characterized by having a temporary storage means given to two or more alphabetic characters and these alphabetic characters which are extracted from the above-mentioned image to memorize temporarily correspondence with the method of two or more kinds of pronunciation as a dictionary.

[Claim 4] It is robot equipment according to claim 2 which is equipped with a word information storage means by which word information including the phonogram and word attribute of a word and this word was memorized as a word attribute table, and is characterized by making the above-mentioned word attribute correspond and memorizing it in the above-mentioned dictionary for speech recognition with the method of the pronunciation of an alphabetic character and this alphabetic character which memorizes the above-mentioned storage control means newly.

[Claim 5] It is robot equipment according to claim 4 which is equipped with the dialogue management tool which generates the response to the voice recognized in the above-mentioned speech recognition means, and is characterized by using the above-mentioned word attribute for the above-mentioned dialogue management tool under the response regulation over

voice.

[Claim 6] The above-mentioned speech recognition means is robot equipment according to claim 2 characterized by recognizing voice based on a hidden Markov model method.

[Claim 7] A storage means for speech recognition by which correspondence relation with the method of the pronunciation of a word and this word was memorized as a dictionary for speech recognition, A word phonetic storage means by which correspondence relation with the phonogram of a word and this word was memorized as a word phonetic table, An image pick-up means to picturize a photographic subject, and an image recognition means to extract the image of a predetermined pattern from the image picturized in the above-mentioned image pick-up means, A sound-collecting means to acquire a surrounding sound, and a speech recognition means to recognize voice from the sound acquired in the above-mentioned sound-collecting means, Two or more kinds of phonograms presumed from the image of the above-mentioned predetermined pattern extracted in the above-mentioned image recognition means are given based on the above-mentioned word phonetic table. A pronunciation information generation means to generate the method of pronunciation, and the voice wave equivalent to pronunciation to each of two or more kinds of phonograms by which grant was carried out [above-mentioned],

The voice wave of the voice recognized in each voice wave and the above-mentioned speech recognition means which were generated in the above-mentioned pronunciation information generation means is compared. The character reader characterized by having a storage control means to memorize newly in the above-mentioned dictionary for speech recognition noting that it is the method of the pronunciation of the alphabetic character which carried out [above-mentioned] the extract of the nearest voice wave.

[Claim 8] The image of the above-mentioned predetermined pattern is a character reader according to claim 7 characterized by being the character string which consists of an alphabetic character and/or two or more alphabetic characters.

[Claim 9] The character reader according to claim 7 characterized by having a temporary storage means given to two or more alphabetic characters and these alphabetic characters which are extracted from the above-mentioned image to memorize temporarily correspondence with the method of two or more kinds of pronunciation as a dictionary.

[Claim 10] It is the character reader according to claim 7 which is equipped with a word information storage means by which word information including the phonogram and word attribute of a word and this word was memorized as a word attribute table, and is characterized by making the above-mentioned word

attribute correspond and memorizing it in the above-mentioned dictionary for speech recognition with the method of the pronunciation of an alphabetic character and this alphabetic character which memorizes the above-mentioned storage control means newly.

[Claim 11] It is the character reader according to claim 10 which is equipped with the dialogue management tool which generates the response to the voice recognized in the above-mentioned speech recognition means, and is characterized by using the above-mentioned word attribute for the above-mentioned dialogue management tool under the response regulation over voice.

[Claim 12] The above-mentioned speech recognition means is a character reader according to claim 7 characterized by recognizing voice based on a hidden Markov model method.

[Claim 13] The image pick-up process which picturizes a photographic subject, and the image recognition process which extracts the image of a predetermined pattern from the image picturized in the above-mentioned image pick-up process, The sound-collecting process which acquires a surrounding sound, and the speech recognition process which recognizes voice from the sound acquired in the above-mentioned sound-collecting process, Two or more kinds of phonograms presumed from the image of the predetermined pattern extracted in

the above-mentioned image recognition process are given based on the word phonetic table on which correspondence relation with the phonogram of a word and this word was memorized. The pronunciation information generation process which generates the method of pronunciation, and the voice wave equivalent to pronunciation to each of two or more kinds of phonograms by which grant was carried out [above-mentioned], The voice wave of the voice recognized in each voice wave and the above-mentioned speech recognition process which were generated in the above-mentioned pronunciation information generation process is compared. The character recognition approach characterized by equipping the dictionary for speech recognition which memorized correspondence relation with the method of the pronunciation of a word and this word for being the method of the pronunciation of the alphabetic character which carried out [above-mentioned] the extract of the nearest voice wave with the storage control process memorized newly.

[Claim 14] The image of the above-mentioned predetermined pattern is the character recognition approach according to claim 13 characterized by being the character string which consists of an alphabetic character and/or two or more alphabetic characters.

[Claim 15] The character recognition approach according to claim 14 characterized by having the process which is given to two or more alphabetic

characters and these alphabetic characters which are extracted from the above-mentioned image, and which is memorized for a temporary storage means temporarily by using correspondence with the method of two or more kinds of pronunciation as a dictionary.

[Claim 16] The character recognition approach according to claim 14 characterized by what a word attribute is made to correspond and is memorized in the above-mentioned dictionary for speech recognition with the method of the pronunciation of the alphabetic character memorized newly and this alphabetic character at the above-mentioned storage control process.

[Claim 17] The character recognition approach according to claim 16 which is equipped with the dialogue management process which generates the response to the voice recognized in the above-mentioned speech recognition process, and is characterized by using the above-mentioned word attribute under the response regulation over voice at the above-mentioned dialogue management process.

[Claim 18] The character recognition approach according to claim 14 characterized by carrying out speech recognition based on a hidden Markov model method at the above-mentioned speech recognition process.

[Claim 19] In the control program of the robot equipment which operates autonomously according to an internal state The image pick-up processing

which picturizes a photographic subject, and the image recognition processing which extracts the image of a predetermined pattern from the image picturized by the above-mentioned image pick-up processing, The sound-collecting processing which acquires a surrounding sound, and the speech recognition processing which recognizes voice from the sound acquired by the above-mentioned sound-collecting processing, Two or more kinds of phonograms presumed from the image of the predetermined pattern extracted by the above-mentioned image recognition processing are given based on the word phonetic table on which correspondence relation with the phonogram of a word and this word was memorized. The pronunciation information generation processing which generates the method of pronunciation, and the voice wave equivalent to pronunciation to each of two or more kinds of phonograms by which grant was carried out [above-mentioned], The voice wave of the voice recognized in each voice wave and the above-mentioned speech recognition processing which were generated by the above-mentioned pronunciation information generation processing is compared. The control program characterized by making robot equipment perform storage processing newly memorized in the dictionary for speech recognition which memorized correspondence relation with the method of the pronunciation of a word and this word for being the method of the pronunciation of the alphabetic character which

carried out [above-mentioned] the extract of the nearest voice wave.

[Claim 20] The image of the above-mentioned predetermined pattern is a control program according to claim 19 characterized by being the character string which consists of an alphabetic character and/or two or more alphabetic characters.

[Claim 21] The image pick-up processing which picturizes a photographic subject, and the image recognition processing which extracts the image of a predetermined pattern from the image picturized by the above-mentioned image pick-up processing, The sound-collecting processing which acquires a surrounding sound, and the speech recognition processing which recognizes voice from the sound acquired by the above-mentioned sound-collecting processing, Two or more kinds of phonograms presumed from the image of the predetermined pattern extracted by the above-mentioned image recognition processing are given based on the word phonetic table on which correspondence relation with the phonogram of a word and this word was memorized. The pronunciation information generation processing which generates the method of pronunciation, and the voice wave equivalent to pronunciation to each of two or more kinds of phonograms by which grant was carried out [above-mentioned], The voice wave of the voice recognized in each voice wave and the above-mentioned speech recognition processing which were generated by the above-mentioned pronunciation information generation

processing is compared. The record medium with which the control program for making robot equipment perform storage processing newly memorized in the dictionary for speech recognition which memorized correspondence relation with the method of the pronunciation of a word and this word for being the method of the pronunciation of the alphabetic character which carried out

[above-mentioned] the extract of the nearest voice wave was recorded.

[Claim 22] The image of the above-mentioned predetermined pattern is a record medium according to claim 21 characterized by being the character string which consists of an alphabetic character and/or two or more alphabetic characters.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] The robot equipment, the character reader, and the character recognition approach this invention operates autonomously according to an internal state, The image of a predetermined pattern is recognized from the image especially picturized about the control program and the record medium in the list. The robot equipment which matches the acquired voice with this

recognition image, and registers it newly with this image, The character reader and the character recognition approach of matching the voice acquired with the image of the picturized predetermined pattern with this recognition image, and registering it into a list newly, It is related with the record medium with which the control program which performs processing which recognizes the image of a predetermined pattern from the acquired image, matches with this recognition image the voice acquired with this image, and registers with a list newly, and this control program were recorded.

[0002]

[Description of the Prior Art] The machinery which performs movement which resembled actuation of human being (living thing) using the electric or magnetic operation is called "robot." Although it was from the end of the 1960s that a robot began to spread in our country, the many were the industrial robots (Industrial Robot) in works aiming at automation, full automation, etc. of production, such as a manipulator and a carrier robot.

[0003] Recently, a life is supported as human being's partner, namely, development of the practical use robot which supports the human activity in various scenes on the everyday life of living conditions and others is furthered. Unlike the industrial robot, such a practical use robot has human being to whom individuality was separately different, or the capacity to learn the adaptation

approach to various environments oneself, in various aspects of affairs of human being's living environment. For example, leg formula mobile robots, such as a "pet mold" robot which imitated the body mechanism of the animal of quadrapedalism and its actuation like a dog and a cat, "a human mold" designed by using as a model the body mechanism of an animal and actuation which perform a 2-pair-of-shoes walk in erect posture, or a "human form" robot (Humanoid Robot), are already being put in practical use. Since these leg formula mobile robots have the appearance configuration possible nearest to the appearance of an animal or human being, actuation near actuation of an animal and human being can be performed as compared with an industrial robot and various actuation which thought entertainment nature as important further can be performed, it may be called an entertainment robot.

[0004] A thing equipped with the miniature camera equivalent to a "eye", the sound-collecting microphone equivalent to a "lug", etc. is also in a leg formula mobile robot. In this case, by performing an image processing to the acquired image, a leg formula mobile robot can recognize the surrounding environment where it inputted as image information, or can recognize "language" from the sound of the inputted perimeter.

[0005] By recognizing the voice especially acquired from the exterior, it changes into an alphabetic character or the technique which recognizes voice, has

answered enough and is carried out is applied to a personal computer and other electronic equipment as a voice recognition unit besides the leg formula mobile robot.

[0006] By the technique of the conventional speech recognition, the pronunciation and the notation of a word are carrying out speech recognition using the dictionary for speech recognition (it is hereafter described as the dictionary for recognition.) matched and memorized. Therefore, there was a fault that it could not recognize about the word which is not registered into the dictionary for recognition. Furthermore, when recognizing the pronunciation of a continuous word like a "sentence", you must be the combination of the word registered into the dictionary for recognition. That is, when the word which is not registered into the dictionary for authentication is contained, it cannot recognize whether it is incorrect-recognized.

[0007] If the word "North Shinagawa" is taken for an example and "North Shinagawa" is not registered into the dictionary for authentication, pronunciation including "North Shinagawa" and "North Shinagawa", for example, the voice which consists of continuation of the word "where North Shinagawa is", cannot be recognized, or the part of "North Shinagawa" is incorrect-recognized. So, in order to enable it to recognize the word which is not registered into the dictionary for recognition, it is necessary to newly carry out additional registration of the

non-registered word.

[0008] The "PLU train" to which the dictionary for recognition which it has in order that a voice recognition unit may make speech recognition possible expresses the pronunciation information on the word as the "word symbol" as an identifier for distinguishing from other words is matched. PLU(s) (Phonone-like unit) are acoustical and a thing used as a phonological unit. The pronounced voice can surely be expressed as a combination (PLU train) of PLU.

[0009] Therefore, what is necessary is just to add a word symbol and the PLU train corresponding to this, when registering a word into the dictionary for recognition. However, the case where a word symbol and a PLU train can be added is restricted when the direct input of the notation "North Shinagawa" and "kitashinagawa" can be carried out using input means, such as a keyboard.

[0010] Therefore, when it does not have an input means like a keyboard like robot equipment, there is also a method of carrying out speech recognition of the pronunciation of the word acquired as voice, and acquiring the PLU train of a strange word. In this case, it recognizes with the application of a garbage model (garbage model). A garbage model is a model (in however, the case of Japanese) which expressed voice as a combination of a "phoneme" used as the fundamental unit of pronunciation, and expressed it as a combination of the "kana" used as the fundamental unit of the reading of a word, as shown in

drawing 20 (a) and drawing 20 (b).

[0011] In the conventional voice recognition unit, by applying a garbage model, the recognition result with voice obtained, applied the word symbol to this recognition result, these were made to correspond, and it has registered with the dictionary for recognition as a new word.

[0012] However, a "phoneme" and "PLU" are mostly used as a word of homonymy, and a "PLU train" writes the pronunciation of the word which consisted of that two or more "PLU(s)" was connected here.

[0013]

[Problem(s) to be Solved by the Invention] however, by the technique of the conventional speech recognition which applied the garbage model A delicate difference being in the method of utterance to every user, even if it is the same word, and weak phonemes (for example, /s/of the initial of the word etc.) There was a fault that becoming that it is hard to be recognized inevitably, change of the phoneme under the effect of a surrounding noise, failure in voice section detection, etc. became a cause, and recognition precision worsened.

[0014] Since it is used in many cases under the situation that the distance of the microphone for the voice acquisition by the side of a voice recognition unit and a user (audio source) is separated when a voice recognition unit is especially applied to robot equipment, the frequency of incorrect recognition becomes high.

[0015] concrete -- for example, -- "-- if the case where cause and ******" is made to recognize is shown -- a recognition result -- "hi to tsu na no ga" and "i tas na ga:" -- like -- "-- although it causes and is similar with ******", it may be recognized as a PLU train which is not the same If speech recognition is performed using the dictionary by which word registration was carried out by such approach, problems, such as a display error by the fall of recognition precision and incorrect recognition, will occur. That is, since the inaccurate PLU train would be given, there was a trouble that the precision at the time of recognizing this word fell in a new registration word.

[0016] the case where a person other than the person who registered pronounces the same word -- temporary -- "-- ****** it caused and ******" was registered into the dictionary for recognition -- from the peculiarity of the pronunciation for every user -- "-- the pronunciation which causes and contains the word ******" may not have been recognized

[0017] Moreover, when changing and displaying the result of speech recognition on an alphabetic character, since the information about a display is not given, the mistaken alphabetic character may be displayed on a new registration word. a "user -- "-- the case where it utters to a voice recognition unit, saying "I want to go to North Shinagawa" after causing and registering ******" with voice -- a voice recognition unit -- even if it causes and ******" is recognized correctly, a display

wants to go" to "hitotsunanoga -- "-- " -- it is one -- " -- I want to go -- " -- it may become Moreover, also when a voice recognition unit repeats the PLU train of a recognition result by speech synthesis, it also produces un-arranging [of being uttered as relation only with the unnatural part of the PLU train of the compounded new registration word].

[0018] Furthermore, the new registration word registered with the garbage model in this way cannot register information about the attribute of words, such as a part of speech and semantics. For example, even if it registers "North Shinagawa", information showing whether this word is a noun or it is the name of a place cannot be registered. therefore -- temporary -- for example, -- a dialogue -- ** -- syntax -- recognition -- ** -- language -- a model -- etc. -- "-- < -- the name of a place -- expressing -- a word -- > -- + -- + -- where -- + -- it is -- + -- it is -- " -- like -- specification -- an expression -- a sake -- the syntax rule -- beforehand -- recording -- having -- **** -- ****** -- new -- registration -- a word -- **** -- being inapplicable -- ** -- saying -- a trouble -- it was . Although it could input with voice also about the attribute of a word at the time of registration, the user needed to know the attribute of a word. Moreover, it is troublesome for a user to input an attribute in addition to the register operation of a word.

[0019] Then, by recognizing the voice which recognizes and acquired the alphabetic character from the picturized image as pronunciation of this

alphabetic character to the voice which this invention is proposed in view of such the conventional actual condition, and is pronounced with the shown alphabetic character As opposed to the voice pronounced with the alphabetic character which the robot equipment which can recognize the new word which could register with the dictionary for recognition by having made the non-registered word into the new word, and was registered further with a sufficient precision, and a list were shown By recognizing the voice which recognizes and acquired the alphabetic character from the picturized image as pronunciation of this alphabetic character The character reader which can recognize the new word which could register with the dictionary for recognition by having made the non-registered word into the new word, and was registered with a sufficient precision, And by picturizing the shown alphabetic character, recognizing an alphabetic character from the picturized image, and recognizing the voice pronounced with presentation as pronunciation of the alphabetic character acquired and recognized It aims at offering the record medium with which the control program which performs processing which registers newly the voice which recognizes and acquired the alphabetic character from the picturized image in the character recognition approach registered into the dictionary for recognition as a new word and the list as pronunciation of this alphabetic character, and this control program were recorded.

[0020]

[Means for Solving the Problem] In order to attain the purpose mentioned above, the robot equipment concerning this invention A storage means for speech recognition by which correspondence relation with the method of the pronunciation of a word and this word was memorized as a dictionary for speech recognition, A word phonetic storage means by which correspondence relation with the phonogram of a word and this word was memorized as a word phonetic table, An image pick-up means to picturize a photographic subject, and an image recognition means to extract the image of a predetermined pattern from the image picturized in the image pick-up means, A sound-collecting means to acquire a surrounding sound, and a speech recognition means to recognize voice from the sound acquired in the sound-collecting means, Two or more kinds of phonograms presumed from the image of the predetermined pattern extracted in the image recognition means are given based on a word phonetic table. A pronunciation information generation means to generate the method of pronunciation, and the voice wave equivalent to pronunciation to each of two or more kinds of given phonograms, The voice wave of the voice recognized in each voice wave and speech recognition means which were generated in the pronunciation information generation means is compared, and the dictionary for speech recognition is equipped with a storage control means to memorize newly

noting that it is the method of the pronunciation of an alphabetic character which extracted the nearest voice wave.

[0021] Such robot equipment gives two or more kinds of phonograms presumed from the image of the predetermined pattern extracted in the image recognition means based on a word phonetic table. The method of pronunciation and the voice wave equivalent to pronunciation are generated to each of two or more kinds of given phonograms. The voice wave of the voice recognized in each voice wave and speech recognition means which were generated in the pronunciation information generation means is compared, and it memorizes newly in the dictionary for speech recognition noting that it is the method of the pronunciation corresponding to the image of the predetermined pattern which extracted the nearest voice wave.

[0022] Especially the image of a predetermined pattern is a character string which consists of an alphabetic character and/or two or more alphabetic characters here.

[0023] Moreover, a storage means for speech recognition by which, as for the character reader concerning this invention, correspondence relation with the method of the pronunciation of a word and this word was memorized as a dictionary for speech recognition; A word phonetic storage means by which correspondence relation with the phonogram of a word and this word was

memorized as a word phonetic table, An image pick-up means to picturize a photographic subject, and an image recognition means to extract the image of a sentence predetermined pattern from the image picturized in the image pick-up means, A sound-collecting means to acquire a surrounding sound, and a speech recognition means to recognize voice from the sound acquired in the sound-collecting means, Two or more kinds of phonograms presumed from the image of the predetermined pattern extracted in the image recognition means are given based on a word phonetic table. A pronunciation information generation means to generate the method of pronunciation, and the voice wave equivalent to pronunciation to each of two or more kinds of given phonograms, The voice wave of the voice recognized in each voice wave and speech recognition means which were generated in the pronunciation information generation means is compared, and the dictionary for speech recognition is equipped with a storage control means to memorize newly noting that it is the method of the pronunciation of an alphabetic character which extracted the nearest voice wave.

[0024] Such a character reader gives two or more kinds of phonograms presumed from the image of the predetermined pattern extracted in the image recognition means based on a word phonetic table. The method of pronunciation and the voice wave equivalent to pronunciation are generated to each of two or

more kinds of given phonograms. The voice wave of the voice recognized in each voice wave and speech recognition means which were generated in the pronunciation information generation means is compared, and it memorizes newly in the dictionary for speech recognition noting that it is the method of the pronunciation of an alphabetic character which extracted the nearest voice wave.

[0025] Especially the image of a predetermined pattern is a character string which consists of an alphabetic character and/or two or more alphabetic characters here.

[0026] Moreover, the image pick-up process at which the character recognition approach concerning this invention picturizes a photographic subject, The image recognition process which extracts the image of a predetermined pattern from the image picturized in the image pick-up process, The sound-collecting process which acquires a surrounding sound, and the speech recognition process which recognizes voice from the sound acquired in the sound-collecting process, Two or more kinds of phonograms presumed from the alphabetic character extracted in the image recognition process are given based on the word phonetic table on which correspondence relation with the phonogram of a word and this word was memorized. The pronunciation information generation process which generates the method of pronunciation, and the voice wave equivalent to pronunciation to each of two or more kinds of given phonograms, The voice wave of the voice

recognized in each voice wave and speech recognition process which were generated in the pronunciation information generation process is compared. The dictionary for speech recognition which memorized correspondence relation with the method of the pronunciation of a word and this word is equipped with the storage control process memorized newly noting that it is the method of the pronunciation of an alphabetic character which extracted the nearest voice wave.

[0027] According to such a character recognition approach, two or more kinds of phonograms presumed from the image of the predetermined pattern extracted in the image recognition process are given based on a word phonetic table. The method of pronunciation and the voice wave equivalent to pronunciation are generated to each of two or more kinds of given phonograms. The voice wave of the voice recognized in each voice wave and speech recognition process which were generated in the pronunciation information generation process is compared, and the dictionary for speech recognition memorizes newly noting that it is the method of the pronunciation of the alphabetic character which the nearest voice wave extracted.

[0028] Especially the image of a predetermined pattern is a character string which consists of an alphabetic character and/or two or more alphabetic characters here.

[0029] Furthermore, the image pick-up processing whose control program

concerning this invention picturizes a photographic subject, The image recognition processing which extracts the image of a predetermined pattern from the image picturized by image pick-up processing, The sound-collecting processing which acquires a surrounding sound, and the speech recognition processing which recognizes voice from the sound acquired by sound-collecting processing, Two or more kinds of phonograms presumed from the alphabetic character extracted by image recognition processing are given based on the word phonetic table on which correspondence relation with the phonogram of a word and this word was memorized. The pronunciation information generation processing which generates the method of pronunciation, and the voice wave equivalent to pronunciation to each of two or more kinds of given phonograms, The voice wave of the voice recognized in each voice wave and speech recognition processing which were generated by pronunciation information generation processing is compared. Robot equipment is made to perform storage processing newly memorized in the dictionary for speech recognition which memorized correspondence relation with the method of the pronunciation of a word and this word noting that it is the method of the pronunciation of an alphabetic character which extracted the nearest voice wave.

[0030] Especially the image of a predetermined pattern is a character string which consists of an alphabetic character and/or two or more alphabetic

characters here. Moreover, an above-mentioned control program is recorded on a record medium, and is offered.

[0031]

[Embodiment of the Invention] The robot equipment shown as an example of 1 configuration of this invention is robot equipment which carries out autonomous working according to an internal state. This robot equipment is a leg formula mobile robot which has an upper extremity, the truncus section, and the membrum inferius at least, and makes a migration means only an upper extremity and the membrum inferius, or the membrum inferius. Although there is robot equipment imitating the pet mold robot imitating the body mechanism of the animal of quadrapedalism or its motion, the body mechanism of the animal of the 2-pair-of-shoes walk which uses only the membrum inferius as a migration means, or its motion in a leg formula mobile robot, the robot equipment shown as a gestalt of this operation is a quadrapedalism type leg formula mobile robot.

[0032] This robot equipment can act according to internal states (getting angry sadness, joy, pleasure, etc.), and also is a practical use robot which supports the human activity in various scenes on the everyday life of living conditions and others, and is the entertainment robot which can express the fundamental actuation which the animal of quadrapedalism performs.

[0033] Especially this robot equipment is the form which imitated the "dog", and

has a head, idiosoma, the upper extremity section, the membrum-inferius section, the tail section, etc. The part equivalent to the joining segment and joint of each part is equipped with the actuator and potentiometer of the number according to the degree of freedom of movement, and control of a control section can express target actuation.

[0034] This robot equipment is equipped with the various sensors for detecting the operation received from the image pick-up section for acquiring a surrounding situation as image data, the microphone section which acquires surrounding voice, and the exterior etc. As the image pick-up section, a small CCD (Charge-Coupled Device) camera is used.

[0035] The robot equipment shown as a gestalt of this operation is equipped with image recognition equipment and a voice recognition unit, and generates the voice wave which gives two or more kinds of reading kanas which extract the image of a predetermined pattern from the image picturized in the CCD camera, and are presumed from the image of the extracted predetermined pattern, and is equivalent to each of two or more kinds of given reading kanas. As a predetermined pattern of an image here, the image of an alphabetic character (character string), an objective configuration, a profile, a shank, and the body itself etc. is raised. And it is robot equipment newly memorizable in the dictionary for speech recognition noting that it is the method (reading) of the pronunciation

corresponding to the image of the predetermined pattern which compared this voice wave with the voice wave of the voice acquired in the microphone section, and extracted the nearest voice wave.

[0036] Hereafter, the robot equipment shown as an example of 1 configuration of this invention is explained with reference to a drawing. The following explanation explains the case where the predetermined pattern recognized from the acquired image is an alphabetic character (character string) to a detail.

[0037] With the gestalt of this operation, robot equipment 1 is the so-called pet mold robot of the configuration which imitated the "dog", as shown in drawing 1. The leg units 3A, 3B, and 3C and 3D are connected with front and rear, right and left of the idiosoma unit 2, the head unit 4 is connected with the front end section of the idiosoma unit 2, the tail section unit 5 is connected and robot equipment 1 is constituted by the back end section.

[0038] As shown in drawing 2 , the control section 16 formed by connecting CPU (Central Processing Unit)10, DRAM (Dynamic Random Access Memory)11, a flash ROM (Read Only Memory) 12, PC (Personal Computer) card interface circuitry 13, and a digital disposal circuit 14 mutually through an internal bus 15 and the dc-battery 17 as a source of power of this robot equipment 1 are contained by the idiosoma unit 2. Moreover, the angular-velocity sensor 18 and acceleration sensor 19 for detecting the sense of robot equipment 1 and the

acceleration of a motion are contained by the idiosoma unit 2.

[0039] The CCD (Charge Coupled Device) camera 20 for picturizing an external situation to the head unit 4, The touch sensor 21 for detecting the pressure received by "it strokes" and the physical influence of "striking" from a user, The distance robot 22 for measuring the distance to the body located ahead, LED (Light Emitting Diode) (not shown) equivalent to the microphone 23 for collecting alien frequencies, the loudspeaker 24 for outputting voice, such as a cry, and the "eye" of robot equipment 1 etc. is arranged in the predetermined location, respectively. CCD camera 20 can picturize the photographic subject which the head unit 4 tends to turn to with a predetermined field angle.

[0040] Actuators 251-25n and Potentiometers 261-26n for free frequency are arranged by the joint part of each leg unit 3A-3D, the joining segment of each leg unit 3A-3D and the idiosoma unit 2, the joining segment of the head unit 4 and the idiosoma unit 2, and the joining segment of the tail section unit 5 and tail 5A, respectively. Actuators 251-25n have the servo motor as a configuration. Leg unit 3A - 3D are controlled by the drive of a servo motor, and it changes in a target posture or actuation.

[0041] LED and each actuators 251-25n are connected with the digital disposal circuit 14 of the control section 16 through the hubs 271-27n corresponding to various sensor lists, such as these angular-velocity sensor 18, an acceleration

sensor 19, a touch sensor 21, a distance robot 22, a microphone 23, a loudspeaker 24, and each potentiometers 261-26n, respectively, and direct continuation of CCD camera 20 and the dc-battery 17 is carried out to the digital disposal circuit 14, respectively.

[0042] A digital disposal circuit 14 incorporates sensor data, and the image data and voice data which are supplied from each above-mentioned sensor one by one, and carries out sequential storing of these through an internal bus 15 in the predetermined location in DRAM11, respectively. Moreover, a digital disposal circuit 14 incorporates the dc-battery residue data showing the dc-battery residue supplied from a dc-battery 17 with this one by one, and stores this in the predetermined location in DRAM11.

[0043] Thus, each sensor data stored in DRAM11, image data, voice data, and dc-battery residue data are used in case CPU10 performs motion control of the robot equipment 1 concerned.

[0044] CPU10 reads the control program stored in the flash ROM 12 at the time of the first stage when the power source of robot equipment 1 was switched on, and stores it in DRAM11. Or CPU10 reads the semiconductor memory equipment with which the PC Card slot of the idiosoma unit 2 which is not illustrated to drawing 1 was equipped, for example, the control program stored in the so-called memory card 28, through the PC card interface circuitry 13, and

stores it in DRAM11.

[0045] CPU10 judges [as mentioned above] the situation of self and a perimeter, and the existence of the directions from a user, and influence based on each sensor data by which sequential storing is carried out, image data, voice data, and dc-battery residue data in DRAM11 from the digital disposal circuit 14.

[0046] Furthermore, CPU10 opts for the action based on this decision result and the control program stored in DRAM11. By driving the actuator needed out of Actuators 251-25n based on the decision result concerned, CPU10 moves the head unit 4 vertically and horizontally, moves the tail of the tail section unit 5, or drives and walks him around each leg unit 3A thru/or 3D. Moreover, CPU10 generates voice data if needed, and supplies it to a loudspeaker 24 through a digital disposal circuit 14. Moreover, CPU10 generates the signal which directs lighting and putting out lights of above-mentioned LED, and LED is turned on or it switches it off.

[0047] Moreover, CPU10 operates a robot according to the demand from the dialogue Management Department 110 grade which a robot is controlled autonomously as mentioned above, and also is mentioned later.

[0048] By these fundamental configurations, robot equipment 1 acts autonomously according to the situation of self and a perimeter, the directions from a user, and influence.

[0049] Furthermore, robot equipment 1 equips the control section 16 of the idiosoma unit 2 with the image speech recognition section 100 as a configuration for registering with the dictionary for speech recognition by making correspondence with the alphabetic character recognized to be the recognized pronunciation into a new registration word. The image speech recognition section 100 has the dialogue Management Department 110, the speech recognition section 120, the output generation section 130, the image-processing character recognition section 140, and the pronunciation information generation section 150, as shown in drawing 3 . As it is indicated in drawing 4 as the dictionary for speech recognition, it is the table which recorded the "PLU train" which expresses the pronunciation information corresponding to this word as the "word symbol" as an identifier for distinguishing from other words. By referring to this dictionary, the notation of the method (reading) of the pronunciation of a word or the word corresponding to pronunciation can be extracted.

[0050] Concretely, the dialogue Management Department 110 generates the response to the voice inputted from utterance of the user who inputted from the microphone 23, dialogue hysteresis, etc. The dialogue Management Department 110 generates the response pattern to the inputted voice based on the various dialogue regulations memorized by the dialogue regulation table 111.

[0051] The speech recognition section 120 changes a user's utterance into the frame for the format which can be processed at the dialogue Management Department 110, for example, text format, syntax analysis, and a dialogue etc. Specifically, the speech recognition section 120 consists of the dictionary 121 for speech recognition, the sound model 122, a language model 123, and sonagraphy section 124 grade. In the sonagraphy section 124, the extract of characteristic quantity required for recognition is performed with a very small time interval. For example, the energy of the acquired sound signal, the number of zero crossovers, a pitch, frequency characteristics, such variation, etc. are extracted. Linear predictive coding (LPC), a fast Fourier transform (FFT), a band pass filter (BPF), etc. are used for a frequency analysis.

[0052] The speech recognition section 120 determines the word sequence corresponding to the characteristic quantity sequence generated in the sonagraphy section 124 using the sound model 122 and the language model 123. As the recognition technique, a hidden Markov model (it is described as HMM below Hidden Markov Model:.) etc. is used, for example.

[0053] In HMM, it is a state-transition model with state transition probability and a probability density function, and changing a condition, the probability value which outputs a characteristic quantity sequence is accumulated, and likelihood is determined. It is the technique used for matching with the method of the

pronunciation given to the alphabetic character recognized in the method of the pronunciation of the word memorized by using the value of the likelihood as a "score" by the dictionary for speech recognition, and the image-processing character recognition section mentioned later. Transition probability, a probability density function, etc. of HMM are a value which leads, learns like the study fault based on the data for study beforehand, and is prepared.

[0054] A sound model can prepare a phoneme (PLU), syllable, a word, a phrase, a sentence, etc. for every unit. for example, Japanese kana "**" - " -- it is --"- " -- obtaining --"- " -- obtaining --"- " --"- " -- it is --"- " -- coming --"-- when the sound model which makes "**" a unit is used, the words of "yes", no ["no"], "good morning", whether "whether to be ****** now", etc. can be constituted by connecting combining these. A phoneme expresses the pronunciation information on a word and is an acoustical and phonological unit. On these specifications, it is used without distinguishing a phoneme and PLU (Phonone-like unit). The pronounced voice can surely be expressed as a combination (PLU train) of a phoneme (PLU).

[0055] According to HMM, similarity with the characteristic quantity sequence of the voice acquired in the language and the microphone 23 which were constituted in this way is calculable as a score. As information for constituting "language" from a sound model, the language model 123 and the dictionary 121

for speech recognition are used. in the dictionary 121 for speech recognition, it is the dictionary in which the method of connection of the sound model (here -- the single character "**" of a kana -- " -- it is -- " ... etc. is shown.) for constituting each word used as the candidate for recognition was shown as a correspondence table, and the language model 123 shows the regulation of the method of connection between a word and a word.

[0056] In the example shown below, in case a "word" is pronounced on recognition processing, it shows the thing with it more convenient [a unit / treat / as one settlement], and is not necessarily in agreement with a linguistic word. For example, although "North Shinagawa" may be treated as one word in the following examples, this may be treated as the two words "north" and "Shinagawa." Furthermore, it can also treat as one word when pronouncing a "North Shinagawa station" and "where a North Shinagawa station to be."

[0057] Moreover, on these specifications, it uses as mind of the hiragana which wrote the "reading kana", or katakana, the "method of pronunciation" is written using a Roman alphabet or a Roman alphabet, and a notation, and it is equivalent to the "phoneme notation" which can be set linguistically. [the reading of the kanji and an alphabetic word] [the actual pronunciation of a reading kana]

[0058] For example, the case where the sentence "the time of - from the time of

-" is treated is considered. in this case -- first -- "0 (****)" "1(****)" ... the word "24 (**********)", and "time (**)" - " -- since -- " -- the method of connection of a word is determined by the language - "until" being alike, respectively, and being related and referring to the sound model 122.

[0059] next, "(word showing figure)" and a "time" -- " -- since -- " -- the method of connection of each word for constituting a sentence is determined by referring to the language model 123 for each word until ["until"] "(word showing figure)", and a "time."

[0060] By applying HMM using this dictionary 121 for speech recognition, and the language model 123, similarity with the characteristic quantity sequence inputted as the sentences "from 2:00 to 5:00" "from 1:00 to 2:00" can calculate as a score. The sentence which consists of a word sequence which has the highest score in it is outputted as a speech recognition result.

[0061] Count of the score in speech recognition processing may be performed by carrying out comprehensive evaluation of the acoustical score given with the sound model 122, and the linguistic score given with the language model 123.

[0062] A linguistic score is a score given based on the transition probability or the chain probability between n continuous words, for example. Transition probability is the value beforehand calculated statistically from a lot of texts, and calls this transition probability "n g" here.

[0063] In addition, a language model may describe the class (what classified the word according to a certain criteria and attribute) of a word also besides describing a word directly in syntax or n g.

[0064] for example, -- describing the syntax "whether < name of a place >+ is + which is + where +", when the words showing the name of a place are collected and the class name the <name of a place> is conferred upon it **** -- the inside of n g -- "-- < name of a place >+ can also prepare the transition probability of + where." In this case, it is n= 3 and transition probability is P (<name of a place> | is where |) correctly.

[0065] The output generation section 130 changes into actual actuation the response pattern which the dialogue Management Department 110 generated. For example, when the response pattern the dialogue Management Department 110 "utters with + "no" which shakes a neck at right and left" is generated, the output generation section 130 generates the voice wave corresponding to "no", and outputs it from a loudspeaker 24 while it generates the pattern of operation which corresponds for "shaking a neck at right and left" in response and sends it to CPU10.

[0066] The image-processing character recognition section 140 identifies the character string which takes with CCD camera 20 and is contained in a ****** image based on the character-pattern database 141. Image patterns, such as an

alphabetic character of the word of each country, are stored in the character-pattern database 141 a hiragana, katakana, the kanji, the alphabet, notations, and if needed. The image-processing alphabetic character discernment section 140 matches between the input image from CCD camera 20, and the image pattern stored in the character-pattern database 141, and recognizes the character string contained in the input image.

[0067] The pronunciation information generation section 150 generates the pronunciation information corresponding to the character string recognized in the image-processing character recognition section 140, i.e., the reading kana of a character string, and generates the method (reading) of the pronunciation further. For example, the reading [case, " came and carry out and / ******"] kana the character string "north Shinagawa" has been recognized to be from the input image is generated, and the method (reading) of the pronunciation "kitashinagawa" is generated in a PLU train.

[0068] The word reading attribute table 151 is a table which read with the word (character string) and described the group of a kana and an attribute, as shown in drawing 4 . The attribute shows the semantics which a word has like the "name of a place", an "identifier", and an "animal."

[0069] When the character string recognized in the image-processing character recognition section 140 is contained in this table, it is reading from this table and

extracting a kana, and the method (reading) of the pronunciation of that character string can be decided from a reading kana. The word reading attribute table 151 is prepared independently [the dictionary 121 for speech recognition].

[0070] the number of vocabularies of the dictionary for recognition -- the convenience on a recognition rate, precision, or processing -- an upper limit -- it is (for example, 65,536 words) -- on the word reading attribute table 151, a word can be described regardless of those limits. This word reading attribute table 151 can also be diverted from other language resources. For example, the dictionary currently used by the kana kanji conversion program, the morphological analysis program, etc. can also be diverted.

[0071] The alphabetic character reading table 152 is a table on which it read with the alphabetic character and correspondence with a kana was described, as shown in drawing 6 . It reads for every notation, alphabet, or single kanji, and the kana is described. If it reads about all usable alphabetic characters and the kana is described, it can read to the character string of arbitration and the method (reading) of pronunciation can be given from a kana.

[0072] The regulation for reading, when the reading grant table 153 is read only on two tables and a kana cannot be given, and giving a kana, and the regulation for specifying this, when a reading kana cannot be specified are described. For example, there are a regulation about phonetic reading and unification of native

Japanese reading, and prolonged-sound-izing, a regulation of ****, a regulation about a repeat, and a regulation that gives reading to an alphabetic word.

[0073] the regulation specifically concerning prolonged-sound-izing -- "... it takes -- " -- "... obtaining -- it is -- " -- etc. -- "... -" -- "... it is the regulation which is acquired and is changed into -" etc. this regulation -- for example, -- "-- ** -- obtaining -- today -- " -- "-- ** -- it is changed into - ****-." the regulation of **** -- for example, reading of "the exit of Shinagawa" -- "-- when carrying out and generating from association with ****** (Shinagawa)" and "**** (opening)", it is the regulation which makes "****" muddy and is made into "****." Moreover, it is the regulation which reads to repeats, such as "**********", corresponding to the regulation about a repeat, and attaches a kana. Furthermore, the regulations which read to an alphabetic word and give a kana are regulations, like the "e" itself carries out vowel reading of the front vowel to instead of [which is not pronounced], when there is "e" in the end of the word of an alphabetic word. For example, in case the reading kana "Taegu" is given to "take", it is the regulation which gives the reading kana "A" to "a" and only gives the reading kana "KU" to "ke."

[0074] Next, the processing at the time of registering a new word into the dictionary for recognition is concretely explained using drawing 7 .

[0075] First, in step S1, it shifts to the word registration mode for word

registration. the shift to word registration mode -- for example, the "register mode" to which a user emits robot equipment 1 and "language -- memorizing -- " -- etc. -- it shifts to word registration mode by making language into a trigger. In addition, a manual operation button is prepared, and when this manual operation button is pushed, it may be made to shift to word registration mode.

[0076] In step S2, directions of the purport which utters the reading of the word into which a user wants to register the notation of a word to register in addition to the directions and/or presentation of a purport which are shown in front of CCD camera 20 of robot equipment 1 are urged to robot equipment 1 to a user. The case where the contents of directions are displayed on the display which robot equipment 1 may direct with voice, and is not illustrated is sufficient as the directions to a user. Here, the word "North Shinagawa" is explained as an example. The kanji, a kana, a Roman alphabet notation, or a PLU train is also available for the alphabetic character presented by the user. concrete -- robot equipment 1 -- "North Shinagawa" -- "-- it causes and any notations, such as ******", "KITASHINAGAWA", and "kitashinagawa", can be recognized.

[0077] In step S3, robot equipment 1 judges whether it is only alphabetic character presentation or there was any utterance with alphabetic character presentation. Only in alphabetic character presentation, when it progresses to step S4 and there is utterance with alphabetic character presentation, it

progresses to step S8 mentioned later. Only in utterance except it, recognition processing by the garbage model is performed as usual.

[0078] First, the case of only alphabetic character presentation is explained. Only in alphabetic character presentation, in step S4, the image-processing character recognition section 140 in robot equipment 1 carries out character recognition (OCR:Optical Charactor Recognition) of what kind of character string is contained in the image picturized in CCD camera 20 based on the character-pattern database 141. here, the candidate of a character recognition result needs to narrow down the image-processing character recognition section 140 to one -- when there is nothing, it leaves two or more candidates. For example, when the recognition result of "*******" is obtained to the alphabetic character "North Shinagawa", it leaves "******."

[0079] Then, in step S5, the pronunciation information generation section 150 in robot equipment 1 generates the method (reading) of the pronunciation of a character string to the character string obtained as a recognition result of step S4. The detail at the time of generating pronunciation is mentioned later. The method (reading) of pronunciation is given to a character string by pronunciation generation processing. When there are two or more recognized character strings, and/or when the method of two or more pronunciation is possible to one character string, all pronunciation patterns are applied.

[0080] In step S6, robot equipment 1 checks to a user which the method (reading) of the pronunciation to the character string generated as mentioned above should adopt among two or more reading [be / it / the right]. In a general case, the method (reading) of pronunciation asks a question as if "reading is the right in OO." When a user returns the "right" and the response of "yes" etc., it progresses to step S7.

[0081] Moreover, when there are two or more kinds of methods (reading) of pronunciation, a question is asked as if [each] "reading is OO." A user adopts the reading which returned the "right" and the response of "yes" etc., and progresses to step S7.

[0082] When the response of "no" etc. is received from a user (i.e., when right reading does not exist), it returns to processing of step S2 or step S4.

[0083] By the above processing, after deciding reading of a new word, the method (reading) of the pronunciation to the character string which progressed to step S7 and was acquired, and this character string is matched, and it registers with the dictionary for recognition as a new word. In case a new word is added, the recognition result of the shown alphabetic character is used for the word symbol column shown in drawing 4 . The method (reading) of the pronunciation decided in step S6 is described by PLU **** corresponding to this character string. Register mode is ended after registering a new word. Then, the

processing for making the updated dictionary for recognition reflect in speech recognition, for example, the reboot of a speech recognition program etc., is performed.

[0084] The case where the alphabetic character written in step S3 on the other hand while the user presented the alphabetic character is uttered is explained. When alphabetic character presentation has utterance, pronunciation information, such as a PLU train, can be generated with a sufficient precision by using cooperatively the information acquired from both.

[0085] Specifically, two or more reading kanas presumed from two or more alphabetic characters presumed from the result of character recognition and each [these] alphabetic character and the method (reading) of the pronunciation corresponding to each reading kana are generated. Thus, by matching the method (reading) of two or more obtained pronunciation, and utterance from a user acquired in the microphone 23, one reading kana and the method (reading) of pronunciation are specified out of two or more candidates generated as mentioned above.

[0086] When there is utterance with alphabetic character presentation, in step S8, character recognition of the image-processing character recognition section 140 in robot equipment 1 is carried out from the image picturized in CCD camera 20. here, the candidate of a character recognition result needs to narrow down

the image-processing character recognition section 140 to one -- when there is nothing, it leaves two or more candidates.

[0087] Then, in step S9, the pronunciation information generation section 150 in robot equipment 1 generates the reading kana of a character string to the character string obtained as a recognition result of step S8. The method (reading) of pronunciation is given to a character string by pronunciation generation processing. When there are two or more recognized character strings, when two or more reading is possible, all pronunciation patterns are applied to one character string.

[0088] Next, in step S10, the temporary dictionary for recognition is temporarily generated from a character string and the method (reading) of pronunciation. This dictionary is hereafter described as the dictionary for recognition for new words. For example, suppose that the alphabetic character "North Shinagawa" picturized by CCD camera 20 has been recognized by two kinds, "North Shinagawa" and "*******", in the image-processing character recognition section 140. The speech information generation section 150 is read to "North Shinagawa" and "*******", and gives a kana. "North Shinagawa" -- "-- it causes, ******" is given, two kinds, "** vacancies are **", are given to "*******", and it is generated further, the method (reading), i.e., the PLU train, of both pronunciation. ["******** is **" and] The dictionary for recognition for new words in this case is

shown in drawing 8 .

[0089] In step S11, speech recognition is performed to utterance from a user using the dictionary for recognition for new words. Speech recognition here is not a continuous speech recognition but word speech recognition. When the user has spoken before rather than the dictionary for recognition for new words is generated, the utterance is recorded and speech recognition is performed to the sound recording voice. The speech recognition in step S11 is discovering a user's utterance and the acoustical nearest word out of the word registered into the dictionary for recognition for new words. However, in processing of step S11, even if a word symbol is the same, when PLU trains differ, it is regarded as another word.

[0090] in drawing 8 , it is utterance of the user out of three words (two "******" regards it as another word) registered here -- "-- it is causing and discovering the word nearest to ******." As a result, the group of a word symbol and a PLU train can be specified as one.

[0091] If the group of a word symbol and a PLU train is specified out of the dictionary for recognition for new words, this will be registered into the dictionary 121 for speech recognition of normal in step S7. Register mode is ended after registering a new word. Then, the processing for making the updated dictionary for recognition reflect in speech recognition, for example, the reboot of a speech

recognition program etc., is performed.

[0092] By processing shown above, robot equipment 1 can register the word which is not memorized by the dictionary 121 for speech recognition as a new word.

[0093] Generation of the method (reading) of the pronunciation of the character string in step S5 and step S9 which were mentioned above is explained to a detail using drawing 9 .

[0094] First, in step S21, it investigates whether the character string recognized by the image-processing character recognition section 140 consists of only kana alphabetic characters. However, with a kana alphabetic character here, a macron "-", a repeat "**- --", etc. are included other than a hiragana and katakana. When the character string consists of only kana alphabetic characters, let the recognized kana alphabetic character be the reading of the character string in step S22. At this time, the pronunciation of prolonged-sound-izing etc. may be corrected a little.

[0095] On the other hand, when the character string recognized by the image-processing character recognition section 140 contains alphabetic characters other than a kana alphabetic character in step S21, in step S23, it distinguishes whether the character string is contained in the word reading attribute table 151.

[0096] When the character string is contained in the word reading attribute table 151, it reads from the table, a kana is acquired, and the method (reading) of pronunciation is generated further (step S24). Moreover, when the attribute of a word is described by the word reading attribute table 151, an attribute is also acquired to coincidence. About the usage of this attribute, it mentions later.

[0097] When the character string is not contained in the word reading attribute table 151, in step S25, it reads combining the reading grant based on the longest match principle and the division minimum method, and the alphabetic character reading table 152, and the reading grant based on a reading grant regulation, and a kana is acquired.

[0098] They are whether the same thing as an input string can be constituted from combining with the longest match principle and the number-of-partitions minimum method two or more words contained in the word reading attribute table 151, and the approach of trying. as a result if "North Shinagawa" and a "station front" are included even if this is not contained in the word reading attribute table 151 when an input string is "in front of a North Shinagawa station", since the "North Shinagawa station front" can be constituted from such combination -- "-- it causes and the reading" before ********** can be acquired. the direction where a longer word is contained when there are two or more kinds of configuration approaches -- giving priority (longest match principle) -- the

direction which can be constituted from fewer words -- giving priority (the number-of-partitions minimum method) -- it carries out and the configuration approach is chosen.

[0099] Moreover, the reading grant based on the alphabetic character reading table 152 divides a character string for every alphabetic character, and is the approach of reading from the alphabetic character reading table 152 for every alphabetic character, and acquiring a kana which divided. Since two or more reading kanas can be given to one kanji in the case of the kanji, the reading kana as the whole character string becomes the combination of the reading kana of each kanji. Therefore, for example, it is the approach of reducing the number of combination using the regulation of "phonetic reading and native Japanese reading cannot be intermingled easily."

[0100] Then, in step S26, a score or reliability is calculated to the candidate of each reading kana acquired by the above-mentioned all directions method, and a high thing is chosen. This reads to the inputted character string and a kana can be given. The method (reading) of pronunciation is generated from the obtained reading kana.

[0101] After passing through each process of step S22, step S24, and step S26, finally in step S27, the method (reading) of the pronunciation to a reading kana is corrected based on regulations, such as prolonged-sound-izing and ****-izing.

[0102] Here, the word reading attribute table 151 is explained to a detail. Only by newly registering a word into the dictionary 121 for speech recognition, the connection regulation between the words recorded on the language model 123 is inapplicable. For example, even if it carries out additional registration of "North Shinagawa" at the dictionary 121 for speech recognition, the chain probability of the syntax or "North Shinagawa" about "North Shinagawa", and other words etc. is not generated only by it. Therefore, ideally, the method of making the connection regulation of a language model reflect in a new registration word adds syntax, or recalculates a chain probability from text data, and although it is reconstituting a language model, a language model is applicable by the simple approach shown below after new registration.

[0103] First, the class name of a <unknown word> is attached to the word which is not contained in the language model. To the language model, the chain probability of a <unknown word> and other words is described. It considers that a new registration word is a <unknown word>, and it calculates the chain probability of this new registration word and other words from the chain probability of a <unknown word> and other words.

[0104] With a class, a word is classified according to a certain criteria and attribute. For example, it classifies according to semantics, and each is named the <name of a place>, a <family name>, and the <name of a country>, or it

classifies according to a part of speech, and each is named a <noun>, a <verb>, and a <adjective>.

[0105] To a language model, the chain probability between classes and the chain probability of a class and a word are described instead of describing the chain probability between words. When searching for the chain probability between words, it investigates to which class a word belongs, the chain probability about the class corresponding to a degree is searched for, and the chain probability between words is calculated from there.

[0106] A class model is applicable by presuming which class it is a word belonging to also about a new registration word at the time of registration.

[0107] When it is made above, in the model for unknown words, the chain probability of the same value is altogether given to a new registration word. With a class model, it becomes a value which is different to which class it belongs to it. Therefore, generally, the linguistic score about a new registration word turns into a score with more suitable using a class model, and is recognized appropriately as a result.

[0108] Therefore, in the word registration by speech recognition, the difficult class name can input easily conventionally. That is, when the character string (word) obtained by character recognition is contained in the word reading attribute table 151, a class name can be acquired from the attribute column of

this table. In addition, in the example shown in drawing 5 , although only one has described the attribute in the attribute column, two or more these can also be described like "< name of a place >, a <proper noun>, and a <name of the station>." When the class the <name of a place> exists in this case, the classification name which is in agreement with a class name, i.e., the <name of a place>, is adopted out of the <name of a place>, a <proper noun>, and a <name of the station>.

[0109] In character recognition, precision of direction recognized including the information about the chain of an alphabetic character may improve rather than recognizing a single character every. Then, the precision of character recognition can be further improved by using the "word symbol" column of the dictionary for recognition, the "word" column of the word reading attribute table 151, etc. as information about the chain of an alphabetic character.

[0110] Although the above explanation explained the case of character recognition as recognition of the predetermined pattern in an acquisition image The alphabetic character (character string) which recognizes the image of the configuration of a body besides an alphabetic character (character string), a profile, a shank, and the body itself, and corresponds as mentioned above is extracted. The voice wave which gives two or more kinds of reading kanas presumed from the extracted alphabetic character, and is equivalent to each of

two or more kinds of given reading kanas is also generable. In this case, in addition to the fundamental configuration shown in drawing 1 , a required configuration is added if needed.

[0111] Thus, by enabling it to master the method of pronunciation as a predetermined pattern corresponding to various cases besides a character string, robot equipment can express signs that information is acquired and learned from the exterior, and entertainment nature can be improved.

[0112] By the way, the robot equipment 1 shown as a gestalt of this operation is robot equipment which can act autonomously according to an internal state. The software configuration of the control program in robot equipment 1 comes to be shown in drawing 10 . As mentioned above, this control program is beforehand stored in the flash ROM 12, and is read at the time of the powering-on early stages of robot equipment 1.

[0113] In drawing 10 , the device driver layer 30 is located in the lowest layer of a control program, and consists of device driver sets 31 which consist of two or more device drivers. In this case, each device driver is the object allowed to carry out direct access to the hardware used by usual computers, such as CCD camera 20 (drawing 2) and a timer, and processes in response to interruption from corresponding hardware.

[0114] Moreover, the ROBOTIKKU server object 32 With the virtual robot 33

which becomes by the software group which offers the interface for being located in the lowest layer of the device driver layer 30, for example, accessing hardware, such as various above-mentioned sensors and Actuators 251-25n With the power manager 34 who becomes by the software group which manages the change of a power source etc. It consists of a device driver manager 35 who becomes by the software group which manages other various device drivers, and a dither INDO robot 36 which becomes by the software group which manages the device of robot equipment 1.

[0115] The manager object 37 consists of an object manager 38 and a service manager 39. The object manager 38 is a software group which manages starting and termination of each software group contained in the ROBOTIKKU server object 32, the middleware layer 40, and the application layer 41, and a service manager 39 is a software group which manages connection of each object based on the initial entry between each object described by the connection file stored in the memory card 28 (drawing 2).

[0116] The middleware layer 40 is located in the upper layer of the ROBOTIKKU server object 32, and consists of software groups which offer the fundamental function of these robot equipments 1, such as an image processing and speech processing. Moreover, the application layer 41 is located in the upper layer of the middleware layer 40, and consists of software groups for opting for action of

robot equipment 1 based on the processing result processed by each software group which constitutes the middleware layer 40 concerned.

[0117] In addition, the concrete software configuration of the middleware layer 40 and the application layer 41 is shown in drawing 11 , respectively.

[0118] As shown in drawing 11 , the middleware layer 40 For noise detection, The object for temperature detection, the object for brightness detection, the object for scale recognition, the object for distance detection, for posture detection, The recognition system 60 which has input semantics converter module 59 grade in the object for touch sensors, the object for motion detection, and each signal conditioning module 50 for color recognition - 58 lists, It consists of output systems 69 which have the object for posture management, the object for tracking, the object for motion playback, the object for a walk, the object for a fall return, an object for LED lighting, and each signal conditioning module 61 for sound playback - 67 grades in output semantics converter module 68 list.

[0119] Each signal conditioning modules 50-58 of the recognition system 60 incorporate the data with which it corresponds of each sensor data read from DRAM11 (drawing 2) by the virtual robot 33 of the ROBOTIKKU server object 32, or image data and voice data, perform predetermined processing based on the data concerned, and give a processing result to the input semantics converter module 59. Here, the virtual robot 33 is constituted by the

predetermined protocol as a part which carries out transfer or conversion of a signal.

[0120] The input semantics converter module 59 It is based on the processing result given from each [these] signal conditioning modules 50-58. "The fall was detected", ["it is "noisy", hot / "hot" /, and bright", "the ball having been detected", and] The self of "it was stroked", "it having been struck", "the scale of C-E-G having been heard", "the body which moves having been detected", "having detected the obstruction", etc. and a surrounding situation, the command from a user, and influence are recognized, and a recognition result is outputted to the application layer 41.

[0121] The application layer 41 consists of five modules, the behavioral model library 70, the action change module 71, the study module 72, the feeling model 73, and the instinct model 74, as shown in drawing 12 .

[0122] As shown in drawing 13 , "when a dc-battery residue decreases", "when avoiding an obstruction, and expressing feeling", the behavioral model library 70 is made to correspond to the condition item of the shoes "at the time of detecting a ball" etc. chosen beforehand, respectively, and the behavioral model which became independent, respectively "is prepared ["a fall return is carried out" and] in it."

[0123] And the time of a recognition result being given from the input semantics

converter module 59, respectively, as for these behavioral models, The parameter value of the corresponding emotion currently held like the after-mentioned at the feeling model 73 if needed when fixed time amount has passed, after the last recognition result is given, It opts for the action which continues while referring to the parameter value of the corresponding desire currently held at the instinct model 74, respectively, and a decision result is outputted to the action change module 71.

[0124] In the case of the gestalt of this operation, in addition, each behavioral model As the technique of opting for the next action As opposed to the arcs ARC1-ARCn1 which connect between to each node NODE0 - NODEn for to other nodes NODE0 of which - NODEn it changes from one node (condition) NODE0 as shown in drawing 14 - NODEn The algorithm called the finite stochastic automaton determined probable based on the transition probability P1-Pn set up, respectively is used.

[0125] Concretely, each behavioral model is made to correspond to the node NODE0 which forms a self behavioral model, respectively - NODEn, respectively, and has the state transition table 80 as shown in drawing 15 for every these node NODE0 - NODEn.

[0126] In this state transition table 80, the input event (recognition result) made into transition conditions in that node NODE0 - NODEn is listed by the line of an

"input event name" at a priority, and the further conditions about that transition condition are described by the "data name" and the corresponding train in the line of the "data range."

[0127] therefore, in the node NODE100 expressed in the state transition table 80 of drawing 15 When the recognition result of "detecting a ball (BALL)" is given The range of "magnitude (SIZE)" of the ball given with the recognition result concerned is "0 to 1000", When the recognition result of "detecting an obstruction (OBSTACLE)" is given, they have been conditions for that the range of "the distance (DISTANCE)" to the obstruction done with the recognition result concerned is "0 to 100" to change to other nodes.

[0128] Moreover, in this node NODE100, when there is no input of a recognition result, it also sets. The inside of each emotion held at the feeling model 73 and the instinct model 74 which a behavioral model refers to periodically, respectively, and the parameter value of each desire, it was held at the feeling model 73 -- "-- glad (Joy) -- " -- "-- surprised (Surprise) -- " -- or -- "-- feeling sad (Sadness) -- " -- when the range of which parameter value is "50 to 100", it can change to other nodes.

[0129] moreover -- a state transition table 80 -- "-- others, while the node name which can change from the node NODE0 - NODEn in the train of the "transition place node" in the column of transition probability" of NODOHE is listed It is

described by the part where it corresponds in the column of transition probability" of NODOHE, respectively. the transition probability to each of other node NODE0 which can change when all the conditions described by the line of an "input event name", a "data name", and the "range of data" are met - NODEn -- "-- others -- the action which should be outputted in case it changes to the node NODE0 - NODEn -- "-- others -- it is described by the line of "output action" in the column of transition probability" of NODOHE. in addition -- "-- others -- the sum of the probability of each line in the column of transition probability" of NODOHE is 100 [%].

[0130] therefore, in the node NODE100 expressed in the state transition table 80 of drawing 15 for example, when the recognition result that it carries out "detecting a ball (BALL)", and the range of "SIZE (magnitude)" of the ball is "0 to 1000" is given It can change to "a node NODE120 (node 120)" by the probability of "30 [%]", and action of "ACTION1" will be then outputted.

[0131] When they are constituted as a lot of nodes NODE0 described as such [respectively] a state transition table 80 - NODEn(s) are connected, and a recognition result is given from the input semantics converter module 59, each behavioral model opts for the next action probable using the state transition table of the node NODE0 - NODEn, and is made as [output / to the action change module 71 / a decision result].

[0132] The action change module 71 shown in drawing 12 chooses from each behavioral model of the behavioral model library 70 the action outputted from the high behavioral model of the priority beforehand defined among the actions outputted, respectively, and sends out the command (this is hereafter called action command.) of the purport which should perform the action concerned to the output semantics converter module 68 of the middleware layer 40. In addition, in the gestalt of this operation, priority is highly set up for the behavioral model written by the bottom in drawing 13 .

[0133] Moreover, the action change module 71 notifies that the action was completed based on the completion information of action given from the output semantics converter module 68 after the completion of action to the study module 72, the feeling model 73, and the instinct model 74.

[0134] On the other hand, the recognition result of the instruction received as influence from a user, such as the study module 72 "was struck" among the recognition results given from the input semantics converter module 59 and "it having been stroked", is inputted.

[0135] And when the study module 72 "is struck" (scolded) based on the notice from this recognition result and the action change module 71, the manifestation probability of that action is reduced, and when "stroked" (praised), the transition probability to which the corresponding behavioral model in the behavioral model

library 70 corresponds so that the manifestation probability of that action may be raised is changed.

[0136] on the other hand, the feeling model 73 -- "-- glad (Joy) -- " -- "-- feeling sad (Sadness) -- " -- "-- getting angry (Anger) -- " -- "-- surprised (Surprise) -- " -- "dislike (Disgust)" -- and -- "-- afraid (Fear) -- " -- the parameter with which the strength of the emotion is expressed for every emotion is held about a total of six emotions. And the feeling model 73 updates the parameter value of each [these] emotion periodically based on the notice from the specific recognition result to which it is given from the input semantics converter module 59, respectively, such as "it having been struck" and "it having been stroked", and elapsed time and the action change module 71 etc.

[0137] The recognition result to which the feeling model 73 is specifically given from the input semantics converter module 59, The amount of fluctuation of the action and its emotion when being computed by predetermined operation expression based on the elapsed time after updating last time etc. of the robot equipment 1 at that time **E [t], The multiplier which expresses the sensibility of E [t] and its emotion for the parameter value of the current emotion is set to ke.

(1) By the formula, as parameter value [of the emotion in the following period] E [t+1] is computed and this is replaced with parameter value [of the current emotion] E [t], the parameter value of the emotion is updated. Moreover, the

feeling model 73 updates the parameter value of all emotions like this.

[0138]

[Equation 1]

$$E[t+1] = E[t] + ke \times \Delta E[t] \quad \cdots (1)$$

[0139] In addition, it is decided beforehand what the notice from each recognition result or the output semantics converter module 68 has effect of on amount of fluctuation **[of the parameter value of each emotion] E [t]. For example, the recognition result of "having been struck" has big effect on amount of fluctuation **[of the parameter value of the emotion of the "resentment"] E [t], and the recognition result of "having been stroked" has big effect on amount of fluctuation **[of the parameter value of the emotion of "joy"] E [t].

[0140] Here, the notice from the output semantics converter module 68 is the so-called feedback information (the completion information of action) of action, and is the information on the appearance result of action, and the feeling model 73 changes feeling also using such information. this -- for example, the feeling level of the resentment falls by action of "barking" -- like -- they are things. In addition, the notice from the output semantics converter module 68 is inputted

also into the study module 72 mentioned above, and the study module 72 changes the transition probability to which a behavioral model corresponds based on the notice.

[0141] In addition, feedback of an action result may be made with the output (action to which feeling was added) of the action change modulator 71.

[0142] on the other hand, "movement avarice (exercise)", "love avarice (affection)", "appetite (appetite)", and "the curiosity (curiosity) of the instinct model 74" are mutually-independent -- the parameter with which the strength of the desire is expressed for these the desires of every is held about four desires the bottom. And the instinct model 74 updates the parameter value of these desires periodically based on the recognition result to which it is given from the input semantics converter module 59, respectively, the notice from elapsed time and the action change module 71, etc.

[0143] The instinct model 74 specifically about "movement avarice", "love avarice", and "curiosity" The amount of fluctuation of that the desire of the when being computed by predetermined operation expression based on the notice from a recognition result, elapsed time, and the output semantics converter module 68 etc. deltaI [k], The parameter value of the current desire as a multiplier ki showing the sensibility of I [k] and its desire As parameter value [of that desire in the following period] I [k+1] is computed using (2) types with a

predetermined period and this result of an operation is replaced with parameter value [of that current desire] I [k], the parameter value of that desire is updated. Moreover, the instinct model 74 updates the parameter value of each desire except "appetite" like this.

[0144]

[Equation 2]

$$I[k+1] = I[k] + ki \times \Delta I[k] \quad \cdots (2)$$

[0145] In addition, it is decided beforehand what the notice from a recognition result and the output semantics converter module 68 etc. has effect of on amount of fluctuation **[of the parameter value of each desire] I [k], for example, it has effect to amount of fluctuation **[of the parameter value of the "fatigue"] I [k] with the big notice from the output semantics converter module 68.

[0146] In addition, in the gestalt of this operation, it is regulated so that each emotion and the parameter value of each desire (instinct) may be changed in the range from 0 to 100, respectively, and the value of multipliers ke and ki is also set up according to the individual for each [an emotion and] the desire of every.

[0147] On the other hand, abstract action commands, such as it being [which is given from the action change module 71 of the application layer 41 as mentioned

above] "advance", "it being glad", the output semantics converter module 68 of the middleware layer 40 "cries", as shown in drawing 11 , or "tracking (a ball is pursued)", are given to the signal conditioning modules 61-67 with which the output system 69 corresponds.

[0148] And these signal conditioning modules 61-67 The servo command value which should be given to the actuators 251-25n (drawing 2) in order to carry out the action based on the action command concerned, if an action command is given, Or the drive data given to LED of a "eye" are generated. the voice data of the sound outputted from a loudspeaker 24 (drawing 2) -- and -- Sequential sending out of these data is carried out at the actuators 251-25n which correspond through the virtual robot 33 and digital disposal circuit 14 (drawing 2) of the ROBOTIKKU server object 32 one by one, a loudspeaker 24, or LED.

[0149] Thus, robot equipment 1 can perform the situation of self (interior) and a perimeter (exterior), the directions from a user, and autonomous action according to influence based on a control program. Therefore, since character-recognition processing mentioned above performs, the character-recognition processing shown in drawing 7 can perform by making the control program for performing processing which determines the method of the pronunciation of the alphabetic character extracted from the image by character recognition processing based on the voice recognized from the surrounding

sound by speech-recognition processing also to the robot equipment which is not equipped with the program read.

[0150] Such a control program is offered through the record medium recorded in the format which robot equipment can read. As a record medium which records a control program, the record medium (for example, a magnetic tape, a floppy (trademark) disk, a magnetic card) of a magnetic reading method, the record medium (for example, CD-ROM, MO, CD-R, DVD) of an optical reading method, etc. can be considered. Storages, such as semiconductor memory (the so-called memory card (configurations, such as a rectangle mold and a square mold, are not asked.), IC card), are also contained in a record medium. Moreover, a control program may be offered through the so-called Internet etc.

[0151] It is reproduced through the reading driver equipment of dedication, or a personal computer, and these control programs are transmitted and read into robot equipment 1 by the cable or wireless connection. Moreover, robot equipment can also read a control program from these storages directly, when it has drive equipment of miniaturized storages, such as semiconductor memory or an IC card. With robot equipment 1, it can read from a memory card 28.

[0152] In addition, as for this invention, it is needless to say for various modification to be possible in the range which is not limited only to the gestalt of operation mentioned above and does not deviate from the summary of this

invention. With the gestalt of this operation, although the robot equipment of quadrapedalism was explained, robot equipment may be a 2-pair-of-shoes walk, and a migration means is not further limited to leg formula move mode.

[0153] Below, the detail of the humanoid robot equipment shown as a gestalt of another operation of this invention is explained. Signs that humanoid robot equipment 200 was viewed from each of the front and back are shown in drawing 16 and drawing 17 . Furthermore, the joint degree-of-freedom configuration which this humanoid robot equipment 200 possesses is typically shown in drawing 18 .

[0154] As shown in drawing 16 , humanoid robot equipment 200 consists of the truncus sections which connect an upper extremity including two arms and a head 201, the membrum inferius which consists of the two legs which realize migration actuation, and an upper extremity and the membrum inferius.

[0155] The neck joint which supports a head 201 has three degrees of freedom called the neck joint yawing axis 202, the neck joint pitching axis 203, and the neck joint roll axes 204.

[0156] Moreover, each carpus consists of the shoulder-joint pitching axis 208, the shoulder-joint roll axes 209, the overarm yawing axis 210, the elbow-joint pitching axis 211, the forearm yawing axis 212, a wrist joint pitching axis 213, a wrist joint roll ring 214, and a hand part 215. Hand parts 215 are the many joints

and the multi-degree-of-freedom structure containing two or more fingers in fact. However, since there are little the contribution and effect to attitude control or walk control of humanoid robot equipment 200, actuation of a hand part 215 is assumed to be a zero degree of freedom on these specifications. Therefore, each arm presupposes that it has seven degrees of freedom.

[0157] Moreover, the truncus section has three degrees of freedom called the truncus pitching axis 205, the truncus roll axes 206, and the truncus yawing axis 207.

[0158] Moreover, each leg which constitutes the membrum inferius consists of the hip joint yawing axis 216, the hip joint pitching axis 217, the hip joint roll axes 218, the knee-joint pitching axis 219, an ankle joint pitching axis 220, ankle joint roll axes 221, and a foot 222. In this specification, the intersection of the hip joint pitching axis 217 and the hip joint roll axes 218 defines the hip joint location of humanoid robot equipment 200. In fact, although the foot 222 of the body is the structure containing the vola of many joints and many degrees of freedom, the vola of humanoid robot equipment 200 makes it a zero degree of freedom. Therefore, each leg consists of six degrees of freedom.

[0159] If the above is summarized, as the humanoid robot equipment 200 whole, it will have 2= 3+7x2+3+6x32 degree of freedom in total. However, the humanoid robot equipment 200 for entertainment is not necessarily limited to 32 degrees of

freedom. It cannot be overemphasized that a degree of freedom, i.e., the number of joints, can be suitably fluctuated according to a constraint, requirement specification, etc. on a design / work.

[0160] The degree of means is mounted using an actuator in fact each one which humanoid robot equipment 200 which was mentioned above has. As for the request of eliminating an excessive swelling by the exterior and making it approximate in the shape of [human] a natural bodily shape, performing attitude control to the unstable structure called a 2-pair-of-shoes walk to an actuator, it is desirable that it is small and lightweight.

[0161] The control-system configuration of humanoid robot equipment 200 is typically shown in drawing 19 . As shown in this drawing, humanoid robot equipment 200 consists of each device unit 230,240,250 R/L and 260 R/L expressing the human limbs, and a control unit 280 which performs adaptive control for realizing coordination actuation between each device unit (however, each of R and L is a suffix which shows each of the right and the left.). the following -- the same .

[0162] Actuation of the humanoid robot equipment 200 whole is controlled by the control unit 280 in generalization. A control unit 280 consists of circumference circuits 282 including the interface (neither is illustrated) which performs the data of the main control section 281 which consists of main circuit components (not

shown), such as CPU (Central Processing Unit) and memory, and each component of a power circuit or humanoid robot equipment 200, and transfer of a command. Especially the installation of this control unit 280 is not limited. Although carried in the truncus section unit 240 in drawing 19 , you may carry in the head unit 230. Or a control unit 280 is arranged out of humanoid robot equipment 200, and you may make it communicate with the airframe of humanoid robot equipment 200 by the cable or wireless.

[0163] Each joint degree of freedom in the humanoid robot equipment 200 shown in drawing 19 is realized by the actuator corresponding to each. That is, the neck joint yawing axis 202, the neck joint pitch 203, the neck joint yawing-axis actuator A2 expressing each of the neck joint roll axes 204, neck joint pitching-axis actuator A3, and neck joint roll-axes actuator A4 are arranged in the head unit 230.

[0164] Moreover, the CCD (ChargeCoupled Device) camera for picturizing an external situation is formed in the head unit 230, and also the touch sensor for detecting the loudspeaker for outputting the microphone for collecting the distance robot for measuring the distance to the body located ahead and alien frequencies and voice and the pressure received by "it strokes" and the physical influence of "striking" from a user etc. is arranged.

[0165] Moreover, truncus pitching-axis actuator A5 expressing each of the

truncus pitching axis 205, the truncus roll axes 206, and the truncus yawing axis 207, the truncus roll-axes actuator A6, and the truncus yawing-axis actuator A7 are arranged in the truncus section unit 240. Moreover, the truncus section unit 240 is equipped with the dc-battery used as the starting power source of this humanoid robot equipment 200. This dc-battery is constituted by the cell in which charge and discharge are possible.

[0166] Moreover, although subdivided by overarm unit 251 R/L, elbow-joint unit 252 R/L, and forearm unit 253 R/L, arm unit 250 R/L The shoulder-joint pitching-axis actuator A8, shoulder-joint roll-axes actuator A9 which the shoulder-joint pitching axis 8, the shoulder-joint roll axes 209, the overarm yawing axis 210, the elbow-joint pitching axis 211, the forearm yawing axis 212, the wrist joint pitching axis 213, and the wrist joint roll axes 214 express respectively, The overarm yawing-axis actuator A10, the elbow-joint pitching-axis actuator A11, the elbow-joint roll-axes actuator A12, the wrist joint pitching-axis actuator A13, and the wrist joint roll-axes actuator A14 are arranged.

[0167] Moreover, although subdivided by femoral region unit 261 R/L, knee unit 262 R/L, and leg part unit 263 R/L, leg unit 260 R/L The hip joint yawing-axis actuator A16, the hip joint pitching-axis actuator A17 expressing each of the hip joint yawing axis 216, the hip joint pitching axis 217, the hip joint roll axes 218,

the knee-joint pitching axis 219, the ankle joint pitching axis 220, and the ankle joint roll axes 221, The hip joint roll-axes actuator A18, the knee-joint pitching-axis actuator A19, the ankle joint pitching-axis actuator A20, and the ankle joint roll-axes actuator A21 are arranged. the actuator A2 used for each joint, and A3 -- it can constitute from a small AC servo actuator of the type which ... one-chip-ized the **** servo control system by the gear direct attachment type more preferably, and was carried in the motor unit.

[0168] Sub control section 235,245,255 R/L of an actuator drive control section and 265 R/L are arranged for every device unit, such as the head unit 230, the truncus section unit 240, the arm unit 250, and each leg unit 260. furthermore, each leg 260 -- while equipping with the touch-down check sensors 291 and 292 which detect whether the vola of R and L was implanted, the attitude sensor 293 which measures a posture is equipped in the truncus section unit 240.

[0169] The touch-down check sensors 291 and 292 consist of proximity sensors or micro switches etc. which were installed in the vola. Moreover, an attitude sensor 293 is constituted by the combination of an acceleration sensor and a gyroscope sensor.

[0170] the output of the touch-down check sensors 291 and 292 -- during periods of operation, such as a walk and transit, -- setting -- each leg on either side -- the present basis or **** -- it can distinguish whether it is in which

condition. Moreover, the inclination and posture of a truncus part are detectable with the output of an attitude sensor 293.

[0171] The main control section 281 can answer the output of each sensors 291-293, and can amend control objectives dynamically. Accommodative control is performed to each of sub control section 235,245,255 R/L and 265 R/L, and, more specifically, the upper extremity of humanoid robot equipment 200, the truncus, and the membrum inferius can realize the exercise-of-the-whole-body pattern driven in cooperation.

[0172] the exercise of the whole body on the airframe of humanoid robot equipment 200 -- a foot -- while setting up movement, a ZMP (ZeroMoment Point) orbit, truncus movement, upper extremity movement, lumbar part height, etc., the command which directs actuation according to these contents of a setting is transmitted to each sub control section 235,245,255 R/L and 265 R/L. And in each sub control sections 235 and 245 and ..., the receiving command from the main control section 281 is interpreted, and a drive control signal is outputted to each actuator A2, A3, etc. "ZMP" here is a point on the floor line where the moment by the floor reaction force during a walk serves as zero, and a "ZMP orbit" means the locus to which ZMP moves during the walk actuation period of humanoid robot equipment 200.

[0173] These moments act on gravity, inertial force, and a list from a walk

system at a road surface with the acceleration produced in connection with gravity and locomotion at the time of a walk. According to so-called "d'Alembert's principle", they balance with the floor reaction force as reaction from a road surface to a walk system, and the floor-reaction-force moment. As a conclusion of dynamic inference, the point (Zero Moment Point), i.e., "ZMP", that a pitch and the roll-axes moment serve as zero exists in the vola grounding point and side top of the support polygon which a road surface forms, or its inside.

[0174] Many of proposals about posture stability control of a leg formula mobile robot or the fall prevention at the time of a walk are used as a norm of stability distinction of a walk of this ZMP. The 2-pair-of-shoes walk pattern generation based on a ZMP norm can set up the point landing [vola] beforehand, and has the advantage of being easy to take the kinematic constraint of the tip of a foot according to a road surface configuration into consideration. Moreover, since making ZMP into a stability distinction norm means treating not the force but an orbit as desired value on kinematic control, feasibility increases technically. In addition, the point which applies ZMP to the stability distinction norm of a bipedal robot at the conceptual list of ZMP is indicated by Miomir Vukobratovic work "LEGGED LOCOMOTION ROBOTS" (work outside Ichiro Kato "a bipedal robot and an artificial guide peg" (Nikkan Kogyo Shimbun)).

[0175] Generally, robots of a 2-pair-of-shoes walk like a humanoid have a high

center-of-gravity location, and the ZMP stable zone at the time of a walk is narrower than quadrapedalism. Therefore, the problem of the posture fluctuation accompanying change of such a road surface condition is divided in a 2-pair-of-shoes bipedal robot, and becomes important.

[0176] as mentioned above, humanoid robot equipment 200 -- each sub control sections 235 and 245, ..., etc. -- the receiving command from the main control section 281 -- interpreting -- each actuator A2 and A3 -- a drive control signal is outputted to ... and the drive of each unit is controlled. Thereby, humanoid robot equipment 200 is stabilized into a target posture, changes and can be walked with the stable posture.

[0177] Moreover, in the control unit 280 in humanoid robot equipment 200, various sensors, such as an acceleration sensor, a touch sensor, and a touch-down check sensor, and the image information from a CCD camera, the speech information from a microphone, etc. are generalized other than attitude control which was mentioned above, and it is processing. In the control unit 280, although not illustrated, various sensors, such as an acceleration sensor, a gyroscope sensor, a touch sensor, a distance robot, a microphone, and a loudspeaker, each actuator, the CCD camera, and the dc-battery are connected with the main control section 281 through the hub which corresponds respectively.

[0178] The main control section 281 incorporates sensor data, and the image data and voice data which are supplied from each above-mentioned sensor one by one, and carries out sequential storing of these through an internal interface in the predetermined location in DRAM, respectively. Moreover, the main control section 281 incorporates the dc-battery residue data showing the dc-battery residue supplied from a dc-battery one by one, and stores this in the predetermined location in DRAM. Each sensor data stored in DRAM, image data, voice data, and dc-battery residue data are used in case the main control section 281 performs motion control of this humanoid robot equipment 200.

[0179] At the time of the first stage when the power source of humanoid robot equipment 200 was switched on, the main control section 281 reads a control program, and stores this in DRAM. Moreover, the main control section 281 judges the situation of self and a perimeter, the existence of the directions from a user, and influence, etc. from the main control section 281 based on each sensor data by which sequential storing is carried out, image data, voice data, and dc-battery residue data to DRAM as mentioned above. Furthermore, the main control section 281 makes humanoid robot equipment 200 take actions, such as the so-called "gesture" and a "gesture", by making a required actuator drive based on the decision result concerned while opting for action according to a self situation based on the control program stored in this decision result and

DRAM.

[0180] Therefore, humanoid robot equipment 200 judges the situation of self and a perimeter based on a control program, and can act autonomously according to the directions from a user, and influence. Moreover, humanoid robot equipment 200 matches and determines the reading presumed from the alphabetic character from which the method (reading) of the pronunciation of the alphabetic character extracted from the image picturized in the CCD camera was extracted, and the voice collected in the sound-collecting microphone. Therefore, the precision of the speech recognition of humanoid robot equipment 200 improves, and a new word can register with the dictionary for speech recognition.

[0181]

[Effect of the Invention] As explained to the detail above, the robot equipment concerning this invention A storage means for speech recognition by which correspondence relation with the method of the pronunciation of a word and this word was memorized as a dictionary for speech recognition, A word phonetic storage means by which correspondence relation with the phonogram of a word and this word was memorized as a word phonetic table, An image pick-up means to picturize a photographic subject, and an image recognition means to extract the image of a predetermined pattern from the image picturized in the image pick-up means, A sound-collecting means to acquire a surrounding sound,

and a speech recognition means to recognize voice from the sound acquired in the sound-collecting means, Two or more kinds of phonograms presumed from the image of the predetermined pattern extracted in the image recognition means are given based on a word phonetic table. A pronunciation information generation means to generate the method of pronunciation, and the voice wave equivalent to pronunciation to each of two or more kinds of given phonograms, The voice wave of the voice recognized in each voice wave and speech recognition means which were generated in the pronunciation information generation means is compared, and the dictionary for speech recognition is equipped with a storage control means to memorize newly noting that it is the method of the pronunciation of an alphabetic character which extracted the nearest voice wave.

[0182] The robot equipment concerning this invention gives two or more kinds of phonograms presumed from the image of the predetermined pattern extracted from the image picturized in the image pick-up means based on a word phonetic table. The method of pronunciation and the voice wave equivalent to pronunciation are generated to each of two or more kinds of given phonograms. It determines noting that it is the method of the pronunciation of an alphabetic character which compared the voice wave of the voice recognized in each voice wave and speech recognition means which were generated in the pronunciation

information generation means, and extracted the nearest voice wave.

[0183] Therefore, according to the robot equipment concerning this invention, the bad influence by incorrect recognition of the pronunciation which contains weak phonemes (for example, /s/of the initial of the word etc.) especially, change of the input phoneme under the effect of a surrounding noise, failure in voice section detection, etc. is inhibited, and the recognition precision at the time of registering a new word can be improved. Since the method of exact pronunciation can memorize in the dictionary for speech recognition by this, the recognition precision at the time of recognizing the word registered as a new word improves..

[0184] Moreover, the robot equipment concerning this invention is equipped with a word information storage means by which word information including the phonogram and word attribute of a word and this word was memorized as a word attribute table, with the method of the pronunciation of the alphabetic character which a storage control means memorizes newly, and this alphabetic character, makes a word attribute correspond and memorizes it in the dictionary for speech recognition.

[0185] Therefore, while according to the robot equipment concerning this invention it becomes unnecessary for a user to input the word attribute information which is needed when applying the syntax rule, a dialogue regulation,

etc. to the inputted voice and the voice to output and convenience improves, when a user does not know attribute information, un-arranging [that attribute information was not able to be inputted] is improved.

[0186] Moreover, a storage means for speech recognition by which, as for the character reader concerning this invention, correspondence relation with the method of the pronunciation of a word and this word was memorized as a dictionary for speech recognition, A word phonetic storage means by which correspondence relation with the phonogram of a word and this word was memorized as a word phonetic table, An image pick-up means to picturize a photographic subject, and an image recognition means to extract the image of a predetermined pattern from the image picturized in the image pick-up means, A sound-collecting means to acquire a surrounding sound, and a speech recognition means to recognize voice from the sound acquired in the sound-collecting means, Two or more kinds of phonograms presumed from the alphabetic character extracted in the image recognition means are given based on a word phonetic table. A pronunciation information generation means to generate the method of pronunciation, and the voice wave equivalent to pronunciation to each of two or more kinds of given phonograms, The voice wave of the voice recognized in each voice wave and speech recognition means which were generated in the pronunciation information generation means is

compared, and the dictionary for speech recognition is equipped with a storage control means to memorize newly noting that it is the method of the pronunciation of an alphabetic character which extracted the nearest voice wave.

[0187] Therefore, according to the character reader concerning this invention, the bad influence by incorrect recognition of the pronunciation which contains weak phonemes (for example, /s/of the initial of the word etc.) especially, change of the input phoneme under the effect of a surrounding noise, failure in voice section detection, etc. is inhibited, and the recognition precision at the time of registering a new word can be improved. Since the method of exact pronunciation can memorize in the dictionary for speech recognition by this, the recognition precision at the time of recognizing the word registered as a new word improves.

[0188] Moreover, the character reader concerning this invention is equipped with a word information storage means by which word information including the phonogram and word attribute of a word and this word was memorized as a word attribute table, with the method of the pronunciation of the alphabetic character which a storage control means memorizes newly, and this alphabetic character, makes a word attribute correspond and memorizes it in the dictionary for speech recognition.

[0189] Therefore, while according to the character reader concerning this

invention it becomes unnecessary for a user to input the word attribute information which is needed when applying the syntax rule, a dialogue regulation, etc. to the inputted voice and the voice to output and convenience improves, when a user does not know attribute information, un-arranging [that attribute information was not able to be inputted] is improved.

[0190] Moreover, the image pick-up process at which the character recognition approach concerning this invention picturizes a photographic subject, The image recognition process which extracts the image of a predetermined pattern from the image picturized in the image pick-up process, The sound-collecting process which acquires a surrounding sound, and the speech recognition process which recognizes voice from the sound acquired in the sound-collecting process, Two or more kinds of phonograms presumed from the alphabetic character extracted in the image recognition process are given based on the word phonetic table on which correspondence relation with the phonogram of a word and this word was memorized. The pronunciation information generation process which generates the method of pronunciation, and the voice wave equivalent to pronunciation to each of two or more kinds of given phonograms, The voice wave of the voice recognized in each voice wave and speech recognition process which were generated in the pronunciation information generation process is compared. The dictionary for speech recognition which memorized correspondence relation with

the method of the pronunciation of a word and this word is equipped with the storage control process memorized newly noting that it is the method of the pronunciation of an alphabetic character which extracted the nearest voice wave.

[0191] Therefore, according to the character recognition approach concerning this invention, the bad influence by incorrect recognition of the pronunciation which contains weak phonemes (for example, /s/of the initial of the word etc.) especially, change of the input phoneme under the effect of a surrounding noise, failure in voice section detection, etc. is inhibited, and the recognition precision at the time of registering a new word can be improved. Since the method of exact pronunciation can memorize in the dictionary for speech recognition by this, the recognition precision at the time of recognizing the word registered as a new word improves.

[0192] Moreover, according to the character recognition approach concerning this invention, it has a word information storage means by which word information including the phonogram and word attribute of a word and this word was memorized as a word attribute table, and with the method of the pronunciation of the alphabetic character which a storage control means memorizes newly, and this alphabetic character, a word attribute is made to correspond and it memorizes in the dictionary for speech recognition.

[0193] Therefore, while according to the character recognition approach

concerning this invention it becomes unnecessary for a user to input the word attribute information which is needed when applying the syntax rule, a dialogue regulation, etc. to the inputted voice and the voice to output and convenience improves, when a user does not know attribute information, un-arranging [that attribute information was not able to be inputted] is improved.

[0194] Furthermore, the image pick-up processing whose control program concerning this invention picturizes a photographic subject, The image recognition processing which extracts the image of a predetermined pattern from the image picturized by image pick-up processing, The sound-collecting processing which acquires a surrounding sound, and the speech recognition processing which recognizes voice from the sound acquired by sound-collecting processing, Two or more kinds of phonograms presumed from the alphabetic character extracted by image recognition processing are given based on the word phonetic table on which correspondence relation with the phonogram of a word and this word was memorized. The pronunciation information generation processing which generates the method of pronunciation, and the voice wave equivalent to pronunciation to each of two or more kinds of given phonograms, The voice wave of the voice recognized in each voice wave and speech recognition processing which were generated by pronunciation information generation processing is compared. Robot equipment is made to perform

storage processing newly memorized in the dictionary for speech recognition which memorized correspondence relation with the method of the pronunciation of a word and this word noting that it is the method of the pronunciation of an alphabetic character which extracted the nearest voice wave.

[0195] Therefore, according to the control program concerning this invention, the bad influence by incorrect recognition of the pronunciation which contains weak phonemes (for example, /s/of the initial of the word etc.) especially, change of the input phoneme under the effect of a surrounding noise, failure in voice section detection, etc. is inhibited, and the recognition precision of robot equipment at the time of registering a new word improves. Since the method of exact pronunciation can memorize in the dictionary for speech recognition by this, the recognition precision at the time of recognizing the word registered as a new word improves.

[0196] Moreover, by recording an above-mentioned control program on a record medium, and offering it, reading of this record medium is possible and the recognition precision at the time of registering a new word improves to the electronic equipment which has a function as a voice recognition unit equipped with an image recognition means and a speech recognition means. Since the method of exact pronunciation is memorizable by this, the recognition precision at the time of recognizing the word registered as a new word improves.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is the external view showing the appearance of the robot equipment shown as an example of 1 configuration of this invention.

[Drawing 2] It is the block diagram showing the configuration of the robot equipment shown as an example of 1 configuration of this invention.

[Drawing 3] It is the block diagram showing the configuration of the image speech recognition section in the robot equipment shown as an example of 1 configuration of this invention.

[Drawing 4] It is drawing explaining the dictionary for speech recognition of the robot equipment shown as an example of 1 configuration of this invention.

[Drawing 5] It is drawing explaining the word reading attribute table of the robot equipment shown as an example of 1 configuration of this invention.

[Drawing 6] It is drawing explaining the alphabetic character reading table of the robot equipment shown as an example of 1 configuration of this invention.

[Drawing 7] The robot equipment shown as an example of 1 configuration of this invention is a flow chart explaining the processing which registers a new word into the dictionary for speech recognition.

[Drawing 8] It is drawing explaining the dictionary for recognition of the robot

equipment shown as an example of 1 configuration of this invention for new words.

[Drawing 9] It is a flow chart explaining the processing which generates the method (reading) of the pronunciation of the character string which the robot equipment shown as an example of 1 configuration of this invention has recognized.

[Drawing 10] It is the block diagram showing the software configuration of the control program of the robot equipment shown as an example of 1 configuration of this invention.

[Drawing 11] It is the block diagram showing the configuration of a middleware layer among the control programs of the robot equipment shown as an example of 1 configuration of this invention.

[Drawing 12] It is the block diagram showing the configuration of an application layer among the control programs of the robot equipment shown as an example of 1 configuration of this invention.

[Drawing 13] It is the block diagram showing the configuration of a behavioral model library among the control programs of the robot equipment shown as an example of 1 configuration of this invention.

[Drawing 14] It is a mimetic diagram explaining the finite stochastic automaton which is an algorithm for opting for action of the robot equipment shown as an

example of 1 configuration of this invention.

[Drawing 15] It is drawing showing the state-transition conditions for opting for action of the robot equipment shown as an example of 1 configuration of this invention.

[Drawing 16] It is an external view explaining the appearance seen from the front of the humanoid robot equipment shown as an example of 1 configuration of this invention.

[Drawing 17] It is an external view explaining the appearance seen from the back of the humanoid robot equipment shown as an example of 1 configuration of this invention.

[Drawing 18] It is drawing showing typically the degree-of-freedom configuration model of the humanoid robot equipment shown as an example of 1 configuration of this invention.

[Drawing 19] It is drawing explaining the control-system configuration of the humanoid robot equipment shown as an example of 1 configuration of this invention.

[Drawing 20] Drawing 20 (a) is the mimetic diagram showing the conventional speech recognition approach which applied the garbage model which makes a "phoneme" a base unit, and drawing 20 (b) is the mimetic diagram showing the conventional speech recognition approach which applied the garbage model

which makes "kana" a base unit.

[Description of Notations]

1 Robot Equipment, 2 Idiosoma Unit, 3A, 3B and 3C, 3D Leg Unit, 4 A head unit, 5 A tail section unit, 10 CPU, 11 DRAM, 12 A flash ROM, 13 PC card interface circuitry, 14 A digital disposal circuit, 15 An internal bus, 16 Control section, 17 A dc-battery, 18 angular-velocity sensor, 19 An acceleration sensor, 20 CCD camera, 21 A touch sensor, 22 A distance robot, 23 A microphone, 24 Loudspeaker, 251-25n An actuator, 261-26n Potentiometer, 271-27n A hub, 28 memory cards, 100 Image speech recognition section, 110 The dialogue Management Department, a 111 dialogue regulation table, 120 Speech recognition section, The dictionary for 121 speech recognition, 122 A sound model, 123 Language model, The 124 sonagraphy sections, 130 The output generation section, 140 Image-processing character recognition section, 141 A character-pattern database, 150 The pronunciation information generation section, 151 A word reading attribute table, 152 An alphabetic character reading table, 153 A reading grant table, 200 humanoid-robot equipment